Fig. 1a-f

M KLH HtH
205
116
97
66
45
HtH1
HtH2
HtH1
HtH2
A340 nm
0,09
0,06
0,03
0
0
20
40
60
80
fraction (no.)
a
b
c
d
e
f

Fig. 1g-m

HtH2
HtH
HtH1
g
HtH1
h
HtH2
i
anti-KLH1
HtH1
j
anti-KLH2
HtH1
k
anti-KLH1
HtH2
l
anti-KLH2
HtH2
m

Fig. 2a-h

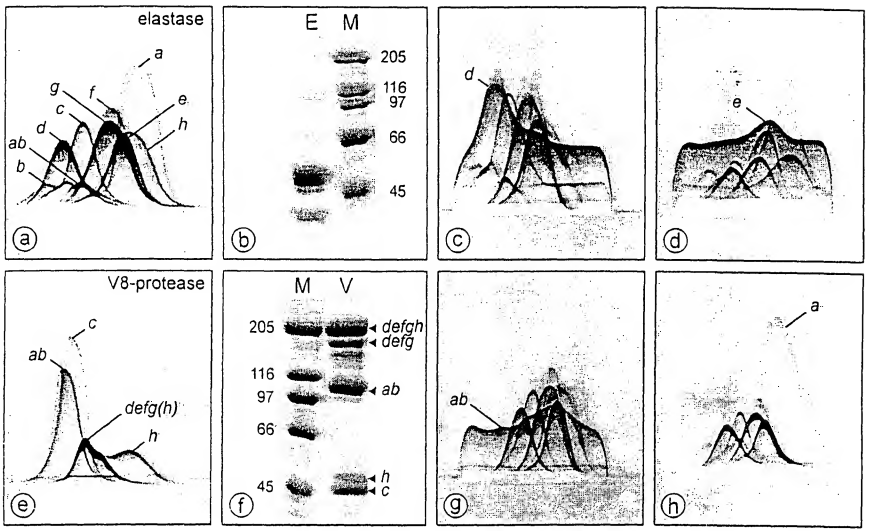

Fig. 2i-p

Fig. 3a-c

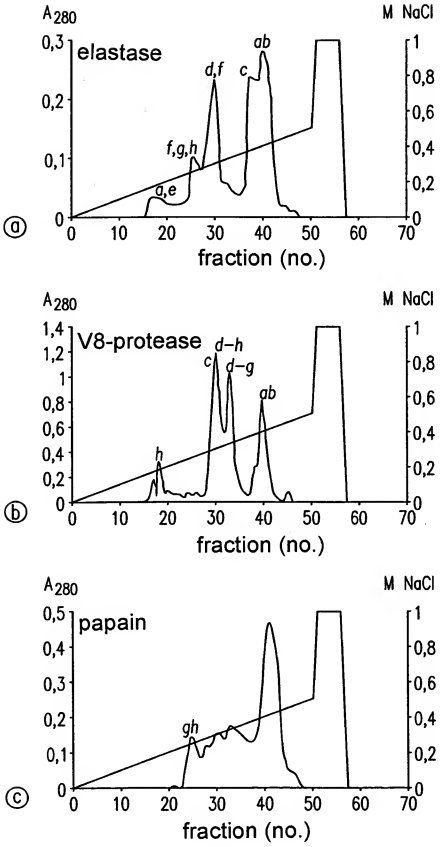
A280
M NaCl
elastase
ab
d,f
c
f,g,h
a,e
fraction (no.)
V8-protease
d–h
c
d–g
ab
h
papain
gh

Fig. 3d-e

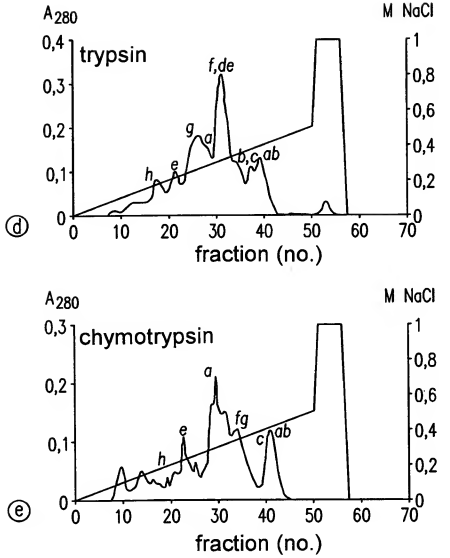
A280
M NaCl
trypsin
f,de
g
a
h
e
b,c,ab
0,4
0,3
0,2
0,1
0
1
0,8
0,6
0,4
0,2
0
0 10 20 30 40 50 60 70
fraction (no.)
ⓓ
A280
M NaCl
chymotrypsin
a
fg
e
h
c
ab
0,3
0,2
0,1
0
1
0,8
0,6
0,4
0,2
0
0 10 20 30 40 50 60 70
fraction (no.)
ⓔ

Figure 4

Genomic sequence of the HtH1 gene

SIGNAL PEPTIDE SEQUENCE 1S-1 (1st part)

GGCTTGTTCAGTTTCTACTCGTCGCCCTTGTG

INTRON 1S-1/1S-2 (SEQ ID NO:109)

GTAAGTCAACGTCTTTGTTTTAAGTTTGATGCATATCTATCATTGCGTTTTAAAATACCA
TTACAACCAACGTGTCTCTATTGGTCTTCACCTGTTTAACGTATATATTGTTTTTAATGT
GAAAATCTGAGATTATTTTCATTTCCGTCAATATTCGTAAAATACTATACAAATAAAATT
GCTTCAGCCTATTGCATTGGCAGTTTTCGCAGAATAACGAGGGAAGGCGTACATAAAATA
TAAACCAGTGTATATTCAAGCATGTTTATAATTTCTTTATAGATTATAACATCATATCAA
AACACCAATCTGGATTTAAACCCGTGAATCCAAAGTATACCAATTAACGGAACTTTATCA
TGTTTTATCAAAGGTTTTAGATGAGGGTAAAGAAGTCCGAGCTATATTTTGCGATATCAG
CAAAGCCTTCTATCACGTCTTGCACACAGGGCTGGTATCTAAACTCGAATCCACAGGAAT
AAATATTTCAGCCGATAGAGAACAGTCGGTGGCTATCATTGGTCACAAAACAAGTCCAAA
ATCTGCATTAGCCGGTGTTCCCCAAGGCTCTGTCTTGGGGCCACTATTATTTCTCACCTA
TATAAACGATTCAACTAATGGAAATATAAAGCAACGTAAACCTCACCGCAGATGAAACACT
AAGTTATAGACAATCCGTTTAAAACCCAGCCACTGCTTAATAATGACTTAGGCCGTCTTT
CAGACTGGGCTAGTAAGCGGCAGGTTAAATTTCACCTTGAAAAGACAGAAACCATGGTAT
ATTTCAAAAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGATGATACTGGGATTT
CTAAAGTGTGTGAACAAAAACACATTGGCCTGATCCTACAAGATAACCAGACAGAAACCA
TGTTTTTTTTCAATAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGATGATACTG
GGATTTCTAAAGTGTGTGGTGAACACAAACACCTTGGCCTGATTCTGCAAGATAATGGAA
AATGTCAGAAACATAAGCAAGTTGATGTGGGGTTTTCTGGGGGTTGTGACAACACCGAAA
GACCCTGCAACTAATGTTAGCTCAAAGGGTTTTACACCCGGTCACAAGTGGGGATCGACC
CAGGCACCTTTTGCCTTTGACAGCTCGCCTTTCAAAAAATCTCAATTCGAAAACGAAATC
TAATAATTTCATGAGCGATACAACCGTTTTTCATAATGCTGTGGTACCGCATACTGTGGA
AACATCTGTCTACCCATTTGGTAGTCCCCCATAAAATGTATTTATGTTTATAAACACAAT
GTTTATAGGGTTACAGTTAGAAGAAGCATTTCTATTGGCTAATGTACATTGCTTGTTTTT
ACTATTGTGCAAAGGCATATTACAGGTCTTTTAGGAAATTAAATACTGTTTAAATCACAT
ACACTACCGGTAATCCTATTATGCTTATCCTGCCAACATTCTGCCCAAGCAAACGCATGA
AAGTTAAAGCTGAGTGTAAAATACTGATTGCTGTGTTACTTCACAACCAGTGGACTGAAT
ACAACCATGTTTTTTCTTGAAAGTCACAAACATCCAGTCGGTTTCTAATGTGTTAAGTTT
CTAGTTTCATAAAGAGCATGACGTAATGGTGAATAGGAGTTATCAATGTTTCTATCTAAT
GACTCCTAGTTCGTTACTTTTTTAATAAAACATCCATGTGTTTAATGTTTGGCCACAGAT
ATAACAAGAAAGAAATCGGATAAAATCTACATTTTGACCAATCGGAAGGCTGCCCCCTCC
CTAATCCTAATCATTTTTGTGCCTCAAAACATACTCAACCAGACATTTGAACTATGTATA
TATCAGAATGAAATGGTAACAATAAACTTGTATGTTGACCAGACAGAATTAGGGTGAATC
TGAATACCAACTATTGTCACATATGAATATGGATAAGCTCTGCGCGTGCGTGCGGGCGGT
GTAGTGCGTGTGTGTCTGTGTGTGTGTGTGTGCGTTTGTGTGTGTCTGCGTGCGTG
TGTGTGCGCGTGTGTGCGTGTGTGTGTGTGCAGTGTGCCGAGTGTGTGTGTGTGTGTG
TGCACAGACATGTGGTTGAGACACACTTGATTCAGTGCAGGATTATGTCCTTCAACCGAG
TGTAGTCTTTAAGTGTGCCTGGAAACAAAAAACTGCGTTGGGTTGCATCGCCTCTGTAGC
AAGCTTGGACGCGTCACGCAGCTCTGATACCACGTATTGGCACCATGTTTCATCGGTCTC
ACGCGAATATTATGCTATGTGTGGCGTATCATACCATAGGTTGGGAACGTTTCAATACTG
TACCGAGCTTGGGCGTGTCACAAAGCTATGATAAGATGACAACACGTCTTGGCATCTTGT
TTCCTCGGTATCACGCGCTGTTATGCTATGTGTGGCTATCACACCTTAGGTTGGGAAAGT

TTCCACATTTTCCAGCCTCGTACATGTTTCCTTTTGTTTTTTCCTTAGTTATCAGCATAC
CGTATATTCTATATTTAATGAGCATTTGTATTTTTCTACAG

SIGNAL PEPTIDE SEQUENCE 1S-2(2nd part)

GTGGGGGCTGGAGCAG

INTRON 1S-2/1A-1 (SEQ ID NO:110)

GTGAGTTTCTTAACATTGTCATGGTACATGGATATACGCTCAGTGGGAAAGCAGGATATC
CCCTTGGTTCAAGTATTCACTTGTCACGCCAAGTGTTCGATTCCCAACATGGAATACTGT
CATATAGTAAATTGATACACTACTTACATTTAATTCTCCACTAAACGTCAACGTCCTTTA
CTTCATGGCCCACATGGTCCGTATTAGTGAGTGAGTGAGTCAGGGCATAAGTATTTAACG
TCAAATCAGCAATATTTCAGCCATATTGTGACAAGAATTGAATATAAATAATTATACTTA
TAATGCTTATAAATATAAATTATATAAATACCTATAACTATAAATTAGTTATACTAGTAT
TTATCAAAACATATTTGCCACGACACTGCACGCCGATACTTCAAGTGTCTTCACCTCAAG
CGTGTAACTCCTCATACTCTGTAATAAGTATGTACACTAAGTGAGTGCTATCATCTCCAT
GCTTCATTAGTTTCGTCAGATGCGTGTATCCATACGAGTACATTCAGATTATGGGATCCA
GAGCTTTCTTATCTCAAGTATTTCCGATTGTAAAGCCATACTACTTCCCCAATGACTGAC
GAGACAGATGGCAACCGTTCTTTCCTCCTGACTAGGTGAGTGCCACTGATAAATCATTAT
GCCTTTAACATTAGGAATGTTAGCAGTGCACATGTTTCAGAATTGCGACCTTATGGTTGT
AAAGATTACAAACTTTACAACTTACTTGAGACAGGTTCCATATGTCGTATCTGAAATAGT
GTGAAGGTATCTGATTCGATGCAATACACAGACATATAAACATATTGTCGCCCTGCTATT
CCGGAAAGGTCATTTTGTATGTAACGTTCCTTAATGGACACAAACGGAATTATTAGTTAA
ACATACTCAACAAAACTATGTTATTTTGCAATGGGTAGCACCGAAATCTACCGACAGTGG
TTCGTAAAAGTAGAACATTCTGACATAAAGAAAAATCATTGGCTTTAAATATATGCAAGT
TACTTGTCTCTAACAACCAGTTTTATACACATTTCAGAGAACGGGGAATCCGCGATGACA
ATATCAACGAGTATATACAGAATATATAATTAAAAACGATGAGTGCCTGGCAAGGGAAAG
AGCGAGATTTGCCAAACAGGGGGGTGGTGTTGAGCTTGAATCGTGGAGAAACGTAGATTG
AAAGACAAGATGACATCTAATGATCCGAAAATCAAACACAGGATTAACTGGGATGCAGAA
GAATGAATATCTCAAGCATACATGCAACACTTCATGAATGCATCTCAAACATTTTCGTCA
GATCGGATGCATGAAGATTTGTAAAGCAATGGTTTAAATTGTCCCTAAACGTTTAGTTGG
AGATGTATGAGGCTAGGCTGTATGTTGAACGAAACCATTTAACATTGTTGTTCATGATTA
TTTAATATTTTTTCATTTTATAGATGTACAATAAAATTGGAAACTAAACATTTCCCTTTA
TTGTTTTGTATTTACCTGTTCATGGGTATGTTTTGAAAGATCGTGATATTTAGTTGGCAT
TCACAAGTTGGAAAAAGGTCACTCAGTTTGATTTCAAGTTTATGTAACCTCTTTATCTGA
CGCTCCAAAATATGTATAGCCTTGTTCATCTGTCGGTATGTGGATATTCCTACTTCAGGG
TAGGGTAGCATTAATACTTACAAAACATAACGTGTACCAGATTTCAGTCACCTCAGAGAT
GATAATGCATGTCGATATGATAGGTCAAAACTTTCGATATCAATCACAATGAACCTATGG
ACCCTGAATCGGAATGATACGTTACACTTTAGAAACAATTCACAAATATGACTGTCACCC
TTTCAGGTAATAATGTTTGACGGACTACGATAGTGCTGAACAGCAGGAGAGGCAACATGG
TTCGATTGTGAGACAGGTTTAGTGTATTTGTTTGCGAATTTAAGGTTCTGAATCACAATA
GACACGGTTCAGTTAATGGATAAACCAATCATTAGATAGATAGAGATTAGTCGCGATATT
GCTGGGATAAAGCTTAGTGGGACGTTAAGTCCCATCTCAATCTCTCTCATTTTTTCCAAA
ACAGTTTTAATTCAGGCTCATGACAAGGTCGTACTGTTGCAAAGGATTCTACTTCAAGCA
GAGATGTCTCATGAATACAGTACAGGGTTTTTGAAGTTTATCCAGTGCAGCGCTGGCACC
ATCTCTGCATGCGAATTATACCATCCATGCCGCTCTAGGCTATTTGTATTAAGTCTGTAG
AATTAAATTCGCGAGTTGCAAATACTGCTCACCATTATCTGCCTCAACCCAGTTTGGGTA
CATGCGATTTACACAATATTATGTATAATGTTCGCTTTTCGAAAACAAAACACCTAAATT
CATCCAAAGTTTTGGGAGATTTTATTCGAGAAATCAACCTGAGATGTTGAATCGGGAGCT
GCGCTTATTCAATGGTGGACTCGGAAGGGAAGTAACCGCTGATGAGGCAAAACAATAACG
CAAACATATGGAAGTGGAACTCTTTGAACCAGTATTATGTTTGTGTGGACATGTATGTGT
TAATTTGACCATTCGAACAACTTTACTATTCTATTCATAATGTGTTTAGATTTACATTTG
AATTAAAAGAGATGAGTTTAAGATATTAATATTTTCCTTTTATAGTCTGTCGTGATTGTA

GGGCAATATTTATGTATGTTCGTTCATTTTTCATTTATCATTTGGAAAGGTATATCATAA
GATTATTATTATCATTCTTGAAGTAATGTATACATATATATATGTCTTGAGTAGCTTATT
TTCAATTTATTATCATCCGTCATCCAATTTTATTTCACGAAAGTATAAGAAATAACGAGA
GAGAGAGAGAGAGAGAGAGAAAAGACAGAAATGAAGTTAGGAGATATNAGTTATCAAGAA
AACAACAGTTTGAATTTTTTGTTTAGACAAGATATCATATCAATAACCTCGCACTATTAC
GGGAATAGGCGGGCGTTCCATATGCACAATGAATCGTCAGTTAAAATCAACATTAAACTT
AAAATACTCCTCATATTTAAAGTTGATCTACCTCTTGTATTATTGTAGACTATTAGACAG
AAGTCGACAGTGACACCAGCAACCAGATATCATACCCAGACTTAAAAAGCTGTTTCCTTG
ATGTTTCAATTTATTTCCATTTCCATTATTTCCCTTTATTGGTTTCCATTTATCAAACTT
ACCATCTGCACCAGTGGGAGATTGATATGTTGTATTTATTTATATTTCTTGTACTACAAT
ATCAAGAATGTATAGGAGCTATTCCTTGTTCCTAAAACCGGATAGATCCATAATTTCCAT
TTTGGGATAAATGGAAACTAAACACAACTTTTACAGTAAACACGAGTGAGCAAGTTGAGT
TTTACGCCGTTTTTAGTAGTATTCCAGCAATATCGCGGCGGGGGACACCAGAAATGGGCT
TCACACAGTGAATGCATGTGGGGATTCGAACCCGGGTCTTCGGCGTGACGAGTGAACGCT
TTAGCCACTAGGCTACCCCACCGCCTATTTATAGTTAAGACGAATACTTTTCTCAAGCCT
CAAATATGTCCATTCTAGAGAGACTGAATCTGATCCTGAATCTGCGGACCGGTCTTGAAT
ATCATCCCACTAACTCATTGTACAAAGTACCTGTAGATTGTCAGTTCAAAGACAGATTTC
ACAACCCTATTATATTTTGTCCTGCTCATTAAGATATTCAGACTCACTCAAACTGCTAAA
TGATTTTAATCCTACTTTGAGATGTTTTAACTTTTATTCGATGCATTTTTGCGTTCTGCG
TCCTGTATAAAGGTAAAGCAGGTAAACTAACCTAACCTGTTGATTTATTTCATAGTTTTG
CGATCAGATTGAAACCGGAATGCACAGTGAAGTGTGGCATACATCTTTCCACAGAGATAC
TGGATACTAGGTGGTACAACCGCATTGGCTTTGTGAAAGGATATTAGTGTTTTATGAGAC
TGACTCATGTTTCAATGCTTAGAGCGGAATGATCTCGGTCTTCATGAAAAATATTGTGTT
GAAGTAACCCCCCAGTCCCTAACAGAACGTGGGGAAAGCAGATGGATATGCCAAGACATC
TTCGCATGGTGTGAAGATGATCGTTACAACATCTGCAGAAAAAGTTATTTCTGTGAAGAA
TATGCCAAAGCATCACTGTGAGTGTTTTGAAGATGTGATATGGCAACACGCAGCGTGTAA
TTATGCTTTGTGTGTATTTCTGAAGATCCGTATGAGCATGGCGCCAAACTATCAGTTAAA
TGGCTATGCGAAGATCTTCCCGAGATGGTAAACACATATTTTGGCCATTTTCTTTGTAAG
TGGGCGACACAGAAGATCCCCCTGATTGTGTGGATGAGGACACAAAAACGGGTCCCCCTT
CCTTTGCTGATGCTAATGACGCCCTGGAAACATTGAAAGACTTCTTCTCCAGCAAGCAAG
CCACCAACCACAAGTTGTATAAATCGCTTGCGGACTTGAATACGGCAGTTGGACAGATAC
ATACAGCCAGAGAGGGCCGAACTAAAACATCTAAACATGGAAAAACTGTAAAGACAGGCT
TTGTTGTACGACGTACGTAAATTCATTGAATGTTTGAAAAGGTAGAAAATTATTAAATCT
TTGAAACCTCGCTCTGTTTGTTTGTTATTGTCCCCCACATTTGCAAATGGTATCCAAAAA
GGGCAGACACATTTGTTTTAATCTTAGCCAGGTTCAATTTAGCCTTGCGCCCAGACTCAT
TGTATCTGGTGAAGGCTATAGGTGGCCACGTCTTCTAAGATGCTATGCTATTCTTACCAG
AATCCAATGTAAAGAGTTCAAACGCATGGTTCGCTTTGATTGTGATTCTTTCTTAGCACC
TCTCTCCTACCCAGAGTTCACCTGCACTGCTCCTGACTCACAATAAGCTGACGTGCTGTC
ATATATGTGCAACATTGTATACGTTGGCGTTAAGCCCAACTCACTTCCGCTGTCTTTTGG
CAG

DOMAIN 1A-1 (1st part of domain a)

ACAACGTCGTCAGAAAGGACGTGAGTCACCTCACAGTTGACGAGGTGCAAGCTCTTCACG
GCGCCCTCCATGACGTCACTGCATCTACAGGGCCTCTGAGTTTCGAAGACATAACATCTT
ACCATGCCGCACCAGCGTCGTGTGACTACAAGGGACGGAAGATCGCCTGCTGTGTCCACG
GTATGCCCAGTTTCCCCTTCTGGCACAGGGCATATGTCGTCCAAGCCGAGCGGGCACTGT
TGTCCAAACGGAAGACTGTCGGAATGCCTTACTGGGACTGGACGCAAACGCTGACTCACT
TACCATCTCTTGTGACTGAACCCATCTACATTGACAGTAAAGGTGGAAAG

INTRON 1A-1/1A-2 (SEQ ID NO:111)

GTAACTACAAACGTCGTCCCATTCATACAGGAGAAATATACAATTGTGTTGTAAGAGCGG
TATACTGTTTGCCAACTGTGTAATTGAAACGTTGATGATGGTGTCTTTGTATTTCAATTT

GTATGCACTTAGACATGATCAATGTTTCTGATGTGTCAAGGATGTTCGGTGTGTCACTTT
CAAAAGATCAAATTCATATGACGTACACAGAGCAAGAACCAACAGTAAGAAGTCTGTATG
ACTTCGCTCTTAAAAGCAATGGAAAAATATTTTCACTTAACACCTAGCCCATAATCACGC
ATATTAGATTATTCAAGCGATGTCAACATGTTTTTAATATCAATCTCATGGTTCTGATAT
TACCGGAGACATGCAACAGGCTGCCATTATAGCCAGGAAATCTTATGAATATGTGCATAT
TTTTTCTTTGATTCTGTATGACGAGAAATATTCGGAGGCAAAGATTGTGTTTTCAGAACA
GAATCAGGGTATCAGTGACATCGTCACTGCATGGCTACAATATTGCTGATGTGACTGTTT
CTCCAAGGATTTTCATCTCACTGTCTGTACTTTGAATCTACAAATTCGTATTAAAGTTAT
GACAATTTTACCCCTGCCTATTTGTAAACGAAATATAACATGAGTGTTTATGCTGACAG

DOMAIN 1A-2 (2nd part of domain a)

GCTCAAACCAACTACTGGTACCGCGGCGAGATAGCGTTCATCAATAAGAAGACTGCGCGA
GCTGTAGATGATCGCCTATTCGAGAAGGTGGAGCCTGGTCACTACACACATCTTATGGAG
ACTGTCCTCGACGCTCTCGAACAGGACGAATTCTGTAAATTTGAAATCCAGTTCGAGTTG
GCTCATAATGCTATCCATTACTTGGTTGGCGGTAAATTTGA

INTRON 1A-2/1A-3 (SEQ ID NO:112)

GTAAGTTTGGTTTACAGTTTCATTATAAAAACATAGCAGTTTTAAGTTTAGGGGCAGATT
CTAATCTCTAATATTCCTTTCAACTCACTTTATTGGTGCCTTCTTGGAGTGACATTTAGA
AACTAAGACAAGAGGAAGATGAACAATGTTTGTAGGGATAGACAGCTTGGATGCAATTTC
GGACCAGATTCTAACAGCGTCATGAAGCAAGTGATACACAACGTTATCAATAACGAGAAT
ATACACATAGATGGTTTGAGTTTATAAATGAACTATTAACGGCATTGTGGTTATAGACAG
TGAGGAAGACGCCAGATAGACAAAGGGTAGGGGCCTTGGTTAGATAATGAGAAGTTGAAG
AGGTGTAATAACTTAAATCTCTCTTGACTATTGATTGTGTCTAAGAGTTTTCTTATCTTA
CAGTCGGCCAGTTGGGTCAAAGATGGTGTGATTCGGATGTGCTTTGTGTGTTCTGCGATG
GCTGATTTAGAGTCAGTTTACTTCAGATGAATGAAGTTCCCCGATTCTTATGTTTAAGTT
TGTTTCACCTACGCATGAAGACATCACCAGCAGGGTCGTCTTTATTTCTAGTAGCTTATT
TACAGCAAGCTTGTAACGTATGCTGAATTGCTGTGCCTCTGTAGAACACAGCATCTATGT
TTGCTTGCTTCTTTAGTAGACTGCGGATGTGATGGTTGGTTACCTGGTATGCTGACGAAA
GAATTGTTGACGTGGTGGTTTGCCTTGATGGGTTCGTTGACTTGGTTTGTTGGATACTGA
TTAAGGTGACTCTGCTGGGAGGCTTGGATTCTGGGGCCGGTGTTCTTTGCTCTCCTGTCT
AGGGTGGCGATTATTTCCCAACCCACTTGTTCCATTACACTCAAAACCTGCTATCAATTT
ACAG

DOMAIN 1A-3 (3rd part of domain a)

ATATTCAATGTCAAACTTGGAATACACCTCCTACGACCCCATCTTCTTCCTCCACCACTC
CAACGTTGACCGCCTCTTCGCCATCTGGCAGCGTCTTCAGGAACTGCGAGGAAAGAATCC
CAATGCAATGGACTGTGCACATGAACTCGCTCACCAGCAACTCCAACCCTTCAACAGGGA
CAGCAATCCAGTCCAGCTCACAAAGGACCACTCGACACCTGCTGACCTCTTTGATTACAA
ACAACTTGGATACAG

INTRON 1A-3/1A-4 (SEQ ID NO:113)

GTGAGACATTATTACACTTCTATTTAGTAGTGGGGGCGGGATAGCTCAGGTGGTAGAGCG
TCGGCCTTCAGCTTCTAGTCTCGCCCACAAGAGCGCGCTGGCTAAAGGGCCGGAGTTAGA
TTCCCGCGGGCGGCAGGCAATATCTCCGAAGGGGAGAACAGTTCTCCAGTCGGTGAAATT
GGGGTGCAATGTTGTACCACTGAAATGCGTGCAGCACCAACCATCCAAATACCAGCCTTG
CCGCGCTGGTCTGACTACATAGTACCACCCGGATTCAACCGGGCTATATAGGTTCTCCTC
CAGCAGTAAATCTGACAGTCGCCATATAGCTGGGATATTGCTGAGTGCGACGTTAAGCCC
CAACTCACTCACTTTATATTTAGTATTCTATTTAGTATCGACGCATGACCATGTGTGGTG
GTCTACTCATCTCAACACGACCGATTAACGTTAAGAGCTGCCAACATGATTCTCTTTCTC

TCTTTAGCCTCTTTATGCCAAAAGCTATATATTAATGTAGGACCCTACATATATTATTTC
CAG

DOMAIN 1A-4 (4th part of domain a)

CTACGACAGCTTAAACCTGAATGGAATGACGCCAGAACAGCTGAAAACAGAACTAGACGA
ACGCCACTCCAAAGAACGTGCGTTTGCAAGCTTCCGACTCAGTGGCTTTGGGGGTTCTGC
CAACGTTGTTGTCTATGCATGTGTCCCTGATGATGATCCACGCAGTGATGACTACTGCGA
GAAAGCAGGCGACTTCTTCATTCTTGGGGGTCAAAGCGAAATGCCGTGGAGATTCTACAG
ACCCTTCTTCTATGATGTAACTGAAGCGGTACATCACCTTGGAGTCCCGCTAAGTGGCCA
CTACTATGTGAAAACAGAACTCTTCAGCGTGAATGGCACAGCACTTTCACCTGATCTTCT
TCCTCAACCAACTGTTGCCTACCGACCTGGGAAAGGTCACCTTGACC

INTRON 1A-4/1B (SEQ ID NO:114)

GTAAGTTGATTGTCTTAATATTGTTTTAATTTTTGCAGAAATTTGATTTTAAATTGTGTA
ATAACAGTACACATTTTTACGCAACAGCAGTCATTATTGTGTGTGAAGATGTCAAACCAG
AAAGGTTTCAATCGTGAAAACAAAAACAATTCTCTATCTGTATACCCCTCAATACCAGTA
TGATCACAAATCTAGGAAATATTACAATACTGCTTCATAGAGTAACTGCTGTTTGTGGCA
GAGCTGGATACGAAGTTTCTGATAGTTCACAGCTACATGATAGTAAATGAACCTGTACAC
ATCAACGGTTGATCATGAAAATTTTGTATGTGTGAAAGTGCTACCTGTATTAGTGAACGT
GCTACCTGTATAACTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGCTGAAA
GTGCTACCTGTATTAGTGAACGTGCTACCTGTATAACTGAAAGTGCTACCTGTATGACTG
AAAGTGCTACCTGTATTAGTGAAAGTGCTACCTGTATGAGTGAACGTGCTACCTGTATAA
CTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATTAGTGAAAGTGCTGCCTGTA
TTAGTGAAAGTGCTACCTGTATGACTGAGCGTGTTACCTGTATGACTGAACGTGCTACCT
GTATTAGTGAAAGTGTAATCTGTATGAGTGAAAGTGCTACCTGTATTAGTGAAAGTGCTA
CTTGTATTAGTGAAAGTGCTACATGTATGACTGAAAGTGCTACATGTATGAATGAGAGTG
CTACCTGTGTGACTGAAAGTGCTACCTGTATTAGTGAAAGTGCTACCTGTATGACTGAAC
GTGCTACCTGTATTAGTGATAGTGTCACTGGTACCAACTGGATGTTCTCACTTCTTTGGC
GAATATCTGGGCTCAAAACAGTTTTTCAGTATCATAGTCGTATCAGTTTGATTTGTATGT
GCAGTGGAATCATTTTCGTCAAATAATCAAAACTGGTGTTGAACTGGCGTTCACGTTTTA
TGGTTGTAAAACAAATTCTGTAAGTAAAGATATTTTAGGGATATCTGTATGACATGAACT
GAATTGCTTAAGGTTAGCATGCCATGACAAATTGCTGAATGTCTGAGGATTGGTGGAGCA
ATAAATCATTATTAAGACAAAAATCAGAAACGTCCATTTTCACTTTTAACAGTGTATCTG
TCTGAATGCCCCCTACTTTTTGGAAGAGTATATATGAATTATCGGCAATATAAAAGCGTTA
AATGGCAAATGTCGGGCATATGTCAGGACATTATTACCGCAGTTTATAGTCATATTTACC
GGGTCTAGGACAATTGTCACCCCGACAATTGCCACCCGGACAATTGCCACCCAAAAATAA
AATATACGTAAACAGAAAACAAATATTGCTTTCAGCCTTTATTGAGTTAGATAATGACAT
TTATGTTGATAAATATGTCGTTTGATAATAATAATAACAATAATATAATATTACAATACT
GCAATAGTACTATCAGTACTTATCATTTTATCACAGATTATATATAGATTCTAGAGTCCG
ATGTTGTAGGCAACACTTCGTCGGTAGGCCGTTAGGTAGTTATCATTAGGGCTGAGTATT
GCGCCAAATTTCGTATTGCTATATACTGCGATACACGGTTACCTGTTTTGCAATACGTAA
ACTTAGGCAAATATGACAGTTTTTCCATGATTATTTTCACGTTTCAATGCTTAAAATGGT
CTTATCTGTTATCTCCTTGAAGGTTTAATAAAATAACAATAAACATAAATCATTATTGAA
AATTAATGAACAAAAGTAAAGCGCTTCTCAGTTACCTTAACCTAACTTATTTATGAATGG
GATTACTATCCAAGAATGTGAAATTCACAAACACCTTGGGATAACACTGCAAAACGACTG
TTCATGGGACGGACATGAAAAAGGTGAGTCCCATGTTAAACTGTTGAGAAAGTTTCCTAT
ACTGTTTGTCCCGAAAAAGGCTAAAGACCATGTACTAATCAATTATTCTATCTATTTTCG
ATTACTGTTCTCATATTTGGGACAACTGTGCAGATCGGTAGCATCCAAGCTCGTCTAAAT
CGGTTTGATAAACCTTGTCAAATAACATGTTGTCTCAACATCCAAGCTCACCTAAACCTT
GTCAATACCTGCATCTGAACAAATGTATATTTAAGACGATAGCATCCAAGCTCATCTTTA
AAATGAATATTTTCTCTTTTTCTACCAAAACATTATTTGGTTGACAGTTGTCCTCCCTAT
TATAGTAAAAAGAACTGGGTGGCAATTGTCCTAGGTGGCAATTGTCCGGATGGCAATTGT

CCGGGTGGCAATTGTCCGGGTGGCAGTTGTCCAGGTGGCTATTGTCCTGTTCCCATATTT
ACGTATCCCATTTTCTGCTCTGTAATTTTAAATAAACTCACCTGCCTAAGGTAAGACGAC
ATGTGTCACGTGAACATCGTTTGGGGGCAAGGGCGGAATCCCTTCGTTGAAAGTAAATGA
ATACTGTACATAGAGATGCGTATCTTGAACTCTTTATTAGCTTTGATATTGTGCTTAATA
TTACATGAATGTATTTCAATATGTAATTATGTGTTCAAATGAATGGTTGACTTGAATGGT
TTTATTGCTTTATATGCTACATCAACATGTGTGTTTCTTTTCATTTCAG

DOMAIN 1B

CACCTGTGCATCATCGCCACGATGACGATCTTATTGTTCGAAAAAATATAGATCATTTGA
CTCGTGAAGAGGAATACGAGCTAAGGATGGCTCTGGAGAGATTCCAGGCCGACACATCCG
TTGATGGGTACCAGGCTACAGTAGAGTACCATGGCCTTCCTGCTCGTTGTCCACGACCAG
ATGCAAAAGTCAGGTTCGCCTGTTGTATGCATGGCATGGCATCCTTCCCTCACTGGCACC
GGCTGTTCGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGGATCGCCTATCGGTGTTC
CTTATTGGGACTGGACAAAACCTATGACTCACCTTCCAGACTTGGCATCAAATGAGACGT
ACGTAGACCCGTATGGACATACACATCATAATCCATTCTTCAATGCAAATATATCTTTTG
AGGAGGGACACCATCACACGAGCAGGATGATAGATTCGAAACTGTTTGCCCCAGTCGCTT
TTGGGGAGCATTCCCATCTGTTTGATGGAATCCTGTACGCATTTGAGCAGGAAGATTTCT
GCGACTTTGAGATTCAGTTTGAGTTAGTCCATAATTCTATTCATGCGTGGATAGGCGGTT
CCGAAGATTACTCCATGGCCACCCTGCATTACACAGCCTTTGACCCCATTTTCTACCTTC
ATCATTCCAATGTCGATCGTCTATGGGCAATCTGGCAAGCTCTTCAAATCAGGAGACACA
AGCCATATCAAGCCCACTGTGCACAGTCTGTGGAACAGTTGCCAATGAAGCCATTTGCTT
TCCCATCACCTCTTAACAACAACGAGAAGACACATAGTCATTCAGTCCCGACTGACATTT
ATGACTACGAGGAAGTGCTGCACTACAGCTACGATGATCTAACGTTTGGTGGGATGAACC
TTGAAGAAATAGAAGAAGCTATACATCTCAGACAACAGCATGAACGAGTCTTCGCGGGAT
TTCTCCTTGCTGGAATAGGAACATCTGCACTTGTTGACATTTTCATAAATAAACCGGGGA
ACCAACCACTCAAAGCTGGAGATATTGCCATTCTTGGTGGTGCCAAGGAAATGCCTTGGG
CGTTTGACCGCTTGTATAAGGTCGAAATAACTGACTCATTGAAGACACTTTCTCTCGATG
TCGATGGAGATTATGAAGTCACTTTTAAAATTCATGATATGCACGGAAACGCTCTTGATA
CGGACCTGATTCCACACGCAGCAGTTGTTTCTGAGCCAGCTCACC

INTRON 1B/1C (SEQ ID NO:115)

GTAAGTAAATTTACAAAATTTGGTGTTCTCTAACTATCCTAAGTATTCAATCGTTAGCGT
GTACCTATCTGCATAATGCAATACCCTGACTCCATATAAGTATAGTATATTTACTCTGGT
CGAAAACAAACAAATTGAAAACAAGAGTGGACGTGCTGTTATGATTTCTTTTTCATTCTT
GGTTCGTTGTGTAATGCCACAGCCAGCAATTCCAGATATATAGCGACGGTCTATGAATAC
TCCAGTCTGGACCAGACAATCGTGTGGAATGGTTTAGGCACATTATATCAAATTCATTGT
TGAAGATATGAGTTATGAGGTCACAATGTTGTCTTGTTACCCCGTGTCAGTAGTGACGTC
ATTTCATGACTGAAATCTCTTCAACGCCGTTTAGCAATAATAGGCTCAGTAGTATTCAAC
CAATTACAATCAGTAGAAAATTCTCTATACTATTCTTATGTTGCATCCTGATATCCCTAT
GCAAAAATTAGTCATCTAATATAATCATTTTCGATAAATACTTTGGGCAAACAAATCAAT
GTAACATCTATTTTCTTTCAG

DOMAIN 1C

CTACCTTTGAGGATGAAAAGCACAGCTTACGAATCAGAAAAAATGTCGACAGCTTGACTC
CTGAAGAAACAAATGAACTGCGTAAAGCCCTGGAGCTTCTTGAAAATGATCATACTGCAG
GTGGATTCAATCAGCTTGGCGCCTTCCATGGAGAGCCTAAATGGTGCCCTAATCCTGAAG
CGGAGCACAAGGTTGCATGCTGTGTTCATGGCATGGCTGTTTTCCCTCATTGGCACAGGC
TTCTTGCTCTCCAGGCGGAGAATGCTCTTAGAAAGCATGGGTACAGTGGTGCTCTACCAT
ACTGGGATTGGACTCGCCCCCTTTCCCAACTTCCTGATCTGGTTAGTCATGAGCAGTATA
CAGATCCTTCCGACCATCACGTGAAGCATAACCCGTGGTTCAATGGCCACATCGATACAG
TAAATCAGGATACCACCAGAAGCGTACGGGAGGATCTTTATCAACAACCTGAATTTGGAC

ATTTCACGGATATTGCTCAACAAGTCCTCTTAGCATTAGAACAAGATGACTTCTGTTCGT
TTGAAGTGCAGTATGAGATTTCCCATAATTTTATCCATGCACTTGTAGGAGGAACCGACG
CTTATGGCATGGCATCGCTGAGATATACAGCATACGATCCAATCTTTTTCTTGCATCATT
CAAACACCGACAGGATCTGGGCTATTTGGCAATCCCTGCAAAAATACAGAGGCAAACCGT
ACAACACTGCCAACTGCGCCATAGAATCTATGAGAAGGCCCCTGCAACCATTTGGACTAA
GCAGTGCCATTAACCCTGACAGAATCACCAGAGAGCATGCTATCCCGTTTGATGTCTTCA
ACTATAGAGATAACCTTCATTACGTATATGATACCCTGGAATTTAATGGTTTGTCGATTT
CACAACTTGATAGAGAGCTGGAAAAAATCAAGAGTCACGAAAGAGTATTTGCTGGATTCT
TGCTGTCGGGGATTAAAAAATCTGCTCTTGTGAAATTCGAAGTTTGTACTCCACCTGATA
ATTGTCATAAAGCAGGGGAGTTTTATCTACTCGGGGACGAAAACGAGATGGCTTGGGCCT
ATGACCGACTTTTCAAGTATGATATTACTCAGGTTCTGGAAGCAAACCATCTACACTTCT
ATGATCATCTCTTCATTCGCTACGAAGTCTTTGATCTTAAAGGAGTGAGTTTGGGAACTG
ACCTGTTCCACACTGCAAATGTGGTACATGATTCCGGCACAG

INTRON 1C/1D (SEQ ID NO:116)

GTACGTGGATTTGATTACATAGCAATGCTATATGATTTCAGTAATTACAACCTCAAGTCA
TGTAGCCGTTTTAGATTGCATTACATCAAACAGCATTGGATTAAATTGGGGGATTGTCCA
GGCCGCATTATGTTGCATTCCGAAAATAGTTTGTGTCCAGTGTCCACGTTTAAAATTAAA
CCATTTTAATCATATTAGGGATAATTTTAATAGATGTTATAGTGCTTTATTTCATATTGT
TACAGTGGACAGTCACCAAGGACATATTTTACTCTATAGATACACAAACACCAATTAAAA
CCCTGCTTTGGAAAGTCTAACTTTTTCCCCACAG

DOMAIN 1D

GCACCCGTGATCGTGATAACTACGTTGAAGAAGTTACTGGGGCCAGTCATATCAGGAAGA
ATTTGAACGACCTCAATACCGGAGAAATGGAAAGCCTTAGAGCTGCTTTCCTGCATATTC
AGGACGACGGAACATATGAATCTATTGCCCAGTACCATGGCAAACCAGGCAAATGTCAAT
TGAATGATCATAATATTGCGTGTTGTGTCCATGGTATGCCTACCTTCCCCCAGTGGCACA
GACTGTATGTGGTTCAGGTGGAGAATGCTCTCCTAAACAGGGGATCTGGTGTGGCTGTTC
CTTACTGGGAGTGGACTGCTCCCATAGACCATCTACCTCATTTCATTGATGATGCCAACAT
ACTTCAATTCCCGACAACAGCGGTACGACCCTAACCCTTTCTTCAGGGGAAAGGTTACTT
TTGAAAACGCAGTCACAACAAGGGACCCACAAGCCGGGCTCTTCAACTCAGATTATATGT
ATGAGAATGTTTTACTTGCACTGGAGCAGGAAAATTATTGTGACTTTGAAATTCAGTTTG
AGCTTGTTCATAACGCACTTCATTCCATGCTGGGAGGTAAAGGGCAGTACTCCATGTCCT
CCCTGGACTATTCTGCGTTTGATCCCGTCTTCTTCCTACATCATGCCAACACGGACAGAC
TGTGGGCAATCTGGCAGGAACTACAAAGATTCCGAGAACTGCCTTATGAAGAAGCGAACT
GTGCAATCAACCTCATGCATCAACCACTGAAGCCGTTCAGTGATCCACATGAGAATCACG
ACAATGTCACTTTGAAATACTCAAAACCACAGGACGGATTCGACTACCAGAACCACTTCG
GATACAAGTATGACAACCTTGAGTTCCATCACTTATCTATCCCAAGTCTTGATGCTACCC
TGAAGCAAAGGAGAAATCACGACAGAGTGTTTGCGGGCTTCCTTCTTCATAACATAGGAA
CTTCTGCTGACATAACTATCTACATATGTCTGCCTGACGGACGGCGTGGCAATGACTGCA
GTCATGAGGCGGGAACATTCTATATCCTCGGAGGCGAAACAGAGATGCCTTTTATCTTTG
ACCGTTTGTATAAATTTGAAATCACCAAACCACTGCAACAGTTAGGAGTCAAGCTGCATG
GTGGAGTTTTCGAACTGGAGCTTGAGATCAAGGCATACAACGGTTCCTATCTGGATCCCC
ATACCTTTGATCCAACTATCATCTTTGAACCTGGAACAG

INTRON 1D/1E (SEQ ID NO:117)

GTAATGCCATCTTAATACAGTTCGTTCGTTAAATTATATATGTTCGTTTACAACACCATA
CCTTGAATTGAGGTAATACATCACTTGATATTGATAATGTAATGGTAATTGTTCTTGTTT
GTAAAACCGTTTCTGGGGTGTTTATTCACTATCCACCTGGTGGATAGTGAGTAAACACAT
TCGGTTTAATATGGGTATCTAATGGACAGTGAAGTGTGCTGGCTAGGCAGATACCTTGGT
TTCTGTGAATGGAGGTAGTAGAAAGGGGTTTTGATGATTGCAG

DOMAIN 1E

ATACCCATATCTTGGACCACGACCATGAGGAAGAGATACTTGTCAGGAAGAATATAATTG
ATTTGAGCCCAAGGGAGAGGGTTTCTCTAGTCAAAGCTTTGCAAAGAATGAAGAATGATC
GCTCCGCTGATGGGTACCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGTCCCA
ATCCATCTGCAGCTCACCGTTATGCTTGCTGTGTCCATGGCATGGCTACATTTCCCCAGT
GGCACAGACTGTACACTGTTCAGGTTCAGGATGCCCTGAGGAGACATGGTTCACTTGTTG
GTATTCCTTACTGGGACTGGACAAAACCAGTCAACGAGTTACCCGAGCTTCTTTCTTCAG
CAACATTTTATCATCCAATCCGGAATATTAATATTTCAAATCCATTCCTCGGGGCTGACA
TAGAATTTGAAGGACCGGGCGTTCATACAGAGAGGCACATAAATACTGAGCGCCTGTTTC
ACAGTGGGGATCATGACGGATACCACAACTGGTTCTTCGAAACTGTTCTCTTTGCTTTGG
AACAGGAAGATTACTGCGATTTTGAAATACAATTTGAGATAGCCCATAATGGCATCCACA
CATGGATTGGTGGAAGCGCAGTATATGGCATGGGACACCTTCACTATGCATCATATGATC
CAATTTTCTACATCCACCATTCACAGACGGACAGAATATGGGCTATTTGGCAAGAGCTGC
AGAAGTACAGGGGTCTATCTGGTTCGGAAGCAAACTGTGCCATTGAACATATGAGAACAC
CCTTGAAGCCTTTCAGCTTTGGGCCACCCTACAATTTGAATAGTCATACGCAAGAATATT
CAAAGCCTGAGGACACGTTTGACTATAAGAAGTTTGGATACAGATATGATAGTCTGGAAT
TGGAGGGGCGATCAATTTCTCGCATTGATGAACTTATCCAGCAGAGACAGGAGAAAGACA
GAACTTTTGCAGGGTTCCTCCTTAAAGGTTTTGGTACATCCGCATCTGTGTCATTGCAAG
TTTGCAGAGTTGATCACACCTGTAAAGATGCGGGCTATTTCACTATTCTGGGAGGATCAG
CCGAAATGCCATGGGCATTCGACAGGCTTTATAAGTATGACATTACTAAAACTCTTCACG
ACATGAACCTGAGGCACGAGGACACTTTCTCTATAGACGTAACTATCACGTCTTACAATG
GAACAGTACTCTCGGGAGACCTCATTCAGACGCCCTCCATTATATTTGTACCTGGACGCC

INTRON 1E/1F-1 (SEQ ID NO:118)

GTGAGTACCTGTTTGCACTAAGACTTCTGTAGGCTAAAAGTGTAAGAAATATCAATTAAT
TTCAATTCACCCAAACTTGAAAACGGTACCTATATAGGTTAACTTTTTGTCTACAGTAAA
CTGAACATACCTACACATTTCATGAAATGATCTCTCAATATTTTCCACCAACAG

DOMAIN 1F-1 (1st part of domain f)

ATAAACTCAACTCACGGAAACATACACCTAACAGAGTCCGCCATGAGCTAAGTAGCCTTA
GTTCCCGTGACATAGCAAGCTTGAAGGCAGCTTTGACAAGCCTTCAACATGATAATGGGA
CTGATGGTTATCAAGCTATTGCTGCCTTCCATGGCGTTCCTGCGCAGTGCCACGAGCCAT
CTGGACGTGAG

INTRON 1F-1/1F-2 (SEQ ID NO:119)

GTAAATTTACAGAGCTTTATGAAGTGTGTTCAGAGTGAAGAGACCAAGATATACTTATAC
CCAAAACTAGCTAGCAACAGACGATTTCACTTGTTTCGGACACTTTGTATTATACGTTGG
ATCCCAAGGTAAACGGAAACGTAACCGAGAATCAGTCCGTAAAGTGAGTGAGTGAGTTTG
GGGCTTAACGTCGCACTCAGCAATACCCCAGCTATGTGGCGACTCTCAGATTTACTGCTG
GAGGAGAACCTACATAGCCCGGTTTAACCCGTGTGGTATGTAGTAAGACCAGCGCGGCAT
GGCTGGTATCTGACGGACGAAGGGTGGCGCTGCACGTATTCCAGTGGTACAACACTGCAC
CCCAATTTCACCGACCGGAGAACTGATCTCCCCTTCGGAGATATCGCCTGCCTTCCACGG
GATTCGAACTCGGTGACCTTCAAGCCAGCGCGCTTCTAGCGGGGGCGATTAGAGGTTNAA
GGCCGACGGCTCTACCACCTTAACTATCCCCCGGCCCCACTCCTGACGGAAATGTTTATA
ATTCAGCCTTTGTTTTCTTATTAAACACTCTTGGCAGATTTTCTATAGATAATGGATTCA
CATGTAGACAGTCTCCCATTGTTGTAACTGGTAGTCAAGAGTTAGAATCTGAATACATTC
TCCAAGATGGATCAAGGAAAACAATAATTACTTGATGTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

ATCGCCTGTTGCATCCACGGCATGGCGACGTTTCCTCACTGGCACCGGTTGTACACTCTG
CAGTTGGAGCAAGCGCTGCGCAGACACGGGTCCAGTGTTGCTGTTCCATACTGGGACTGG
ACCAAGCCAATCACCGAACTGCCACACATTCTGACAGACGGAGAATATTATGACGTTTGG
CAAAATGCCGTCTTGGCCAATCCGTTTGCAAGAGGTTATGTGAAAATTAAAGATGCATTT
ACGGTGAGAAATGTCCAGGAAAGTCTGTTCAAAATGTCAAGTTTTGGAAAGCACTCGCTT
CTGTTTGACCAGGCTTTGTTGGCTCTTGAACAAACTGACTACTGTGACTTCGAAGTTCAG
TTTGAAGTGATGCATAACACGATCCATTATCTCGTAGGAGGGCGTCAAACGTACGCCTTC
TCCTCTCTCGAGTATTCCTCATACGATCCAATCTTCTTTATTCACCACTCGTTTGTTGAC
AAAATATGGGCTGTATGGCAAGAACTGCAAAGCAGGAGACATCTACAGTTTAGAACAGCT
GATTGTGCTGTGGGCCTCATGGGTCAGGCAATGAGGCCTTTCAACAAGGATTTCAACCAC
AACTCGTTCACCAAGAAGCACGCAGTCCCTAATACAGTATTTGATTATGAAGATCTTGGC
TATAACTATGACAACCTTGAAATCAGTGGTTTAAACTTAAATGAGATCGAGGCGTTAATA
GCAAAACGCAAGTCACATGCTAGAGTCTTTGCTGGGTTCCTGTTGTTTGGATTAGGAACT
TCGGCTGATATACATCTGGAAATTTGCAAGACATCGGAAAACTGCCATGATGCTGGTGTG
ATTTTCATCCTTGGAGGTTCTGCAGAGATGCATTGGGCATACAACCGCCTCTACAAGTAT
GACATTACAGAAGCATTGCAGGAATTTGACATCAACCCTGAAGATGTTTTCCATGCTGAT
GAACCATTTTTCCTGAGGCTGTCGGTTGTTGCTGTGAATGGAACTGTCATTCCATCGTCT
CATCTTCACCAGCCAACGATAATCTATGAACCAGGCGAAG

INTRON 1F-2/1G-1 (SEQ ID NO:120)

GTGAGATATATGCAAATTGAATGTTGTCCAGATGCGTTGTTTACATTTATATGCTTGGAA
TTGTCCTGAACGAATACAGTGGAATAACCAAAAGCTGAAAAATAAAAAGATATATACTTC
ATTCTGAATTTGTCAGTATTGCTGACCCAAAAACACGTTATCCATGTCGACACTATATTT
GCCTTTCTGAATCTGAGACTGCGTTATGTTTCTAATAATCACGAAATATGGTATACAGGT
TGTGTATCTGTAGAATACCCAAGGCAGAATTTAAAGGGTCACACCCTGTTTAATACAG

DOMAIN 1G-1 (1st part of domain g)

ATCACCATGACGACCATCAGTCGGGAAGCATAGCAGGATCCGGGGTCCGCAAGGACGTGA
ACACCTTGACTAAGGCTGAGACCGACAACCTGAGGGAGGCGCTGTGGGGTGTCATGGCAG
ACCACGGTCCCAATGGCTTTCAAGCTATTGCTGCTTTCCATGGAAAACCAGCTTTGTGTC
CCATGCCTGATGGCCACAACTACTCATGTTGTACTCACG

INTRON 1G-1/1G-2 (SEQ ID NO:121)

GTAAGTTTGTGTTGGTTAGTGTTGGTTGCATGTTTTGCCATATCGATAGTATCAGTGTGG
TAACATCTGGTTTCTAGTTCATTCAGTTCACCTTATCAGAAGCTGTTTGCTCTCGTCTAC
AATAGTGACGTCTTTCAGTTTTAGAACCGTGTACATCCGGGTTATATTGGTCTCCAGCAA
CCCGTGCTTGTCGTGGGAGGCCACTGATGGGAACGGGTGGTCAGACTCGCTCACTTAGTT
GACACATGTCAATTGCGAAGATCGATGCTGAGGTTGTTAAACATTGGATTGTCTGGTCCA
GACTCGATTATTTACAGACAGCCGCCATGTACCTGGAATATTGCTGAGTGCGGCGTTAAA
CAACAAACTAGTCAGACTAATCTTTCACTGTTTATAATGATGGCTCGAACCTAGCACTCA
TGTCCCAAGTTGGCGAACATCTGGAAGGGAATTTCAAATGAAAAGAACAATCTTTCACGT
CTATTGGTATCACGCTCCTGGAGAAGAACATGATGTTCACGGCGTTACTTCCTCTTACCT
GTTTTACTTGTTCCCACGTTTCTTCATATTTAAAGAGTATTTGGGTATTAGAGCTTTGGT
GCTGTTACAATGCTACTCAACTGTTCAGTGCGGGCGACCGCGCTTGTTTACACATTAAGT
TTTGTTTGTTGGTTGGTTTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTA
TGTGTGTGTGTGTATCTATGTCTATGTGTCTGTGTCTGTGTCTGTCTATGTGTGTG
TGTGTCTGTGTCTATGTGTGTGTCTGCGTGTGTGTCTGTGTCCGTATGTGGCTGTGTCTA
TGTGTGTGTGTGTCTGTGTTTATGTGTGTATATGCGTGTGTGTCTGTGTCCGTATGTGGC

TGTGTCTATGTGTGTGACATGCAATACATGCTGTGATACTCACTAGCTGCGTCTATCGAC
CAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACCTTCCCACACTGGCATCGCCTCTACACCAAGCAGATGGAGGATGCAATGA
GGGCGCATGGGTCTCATGTCGGCCTGCCCTACTGGGACTGGACTGCTGCCTTCACCCACC
TGCCAACACTGGTCACCGACACGGACAACAACCCCTTCCAACAT

INTRON 1G-2/1G-3 (SEQ ID NO:122)

GTAAGAGCGGGGTAGGGATGGGGTGGTAGGGGGTGGGTTGTTCTATTACTTCCCGCTTCA
CTTGTATGAAATGGATAACCTTGGCTGCATCCCAATTGCGTGATCGATTCTCTTTCGATT
CACTCGTGCGATTAGACTGCCTTATTTACTATAGTAGTTAGAATGTTGCTCAGTGCGCCG
TTAAACAACTAATACACAAAACCGCATTTGTTTTATATGGTCACTCTACTGTTTATCACG
TATATGTATGTTCCGACTCACTGGTTGGTGCGTACCATTCTACTGTCACACTGAGAGCCA
ATGTTCTCAGATGTGTGAAATGTTTGAAAGCCGTTTCTACATAATATTGCAGGAATACCA
TTGTAGAATGTAGTCAAACAGGTAACAATCTGTTAGTGAGCCCAGTTCGAGGTTGCGTTG
TAGGGTGTAGTCCAACAGGTAGGCAGTCCATAAGCATAGTTTTTAAGCATTTTAGATCAT
CTATAATTAACCACATGGTTAGCCGCTATGTTTAGTTTAATCCAGTATAAGTTAGAACTG
TTATATTTCGAAGGGAAGTGAGTAAATCCTTATTCCTTGACTACCATTTAATAGATTTCC
CAATGACTCCATTCAACTCCTAACTTTCACATCACTGCTCTCTTCAACAG

DOMAIN 1G-3 (3rd part of domain g)

GGACACATTGATTATCTCAATGTCAGCACAACTCGATCTCCCCGAGACATGCTGTTCAAC
GACCCCGAGCATGGATCAGAGTCGTTCTTCTACAGACAAGTCCTCTTAGCTCTGGAACAA
ACTGATTTCTGCAAATTCGAAGTTCAGTTTGAGATAACCCACAATGCCATCCATTCCTGG
ACAGGTGGCCACAGCCCCTACGGAATGTCCACTCTCGACTTCACTGCCTACGATCCTCTC
TTCTGGCTTCACCACTCCAACACCGACAGAATCTGGGCTGTCTGGCAAGCTTTGCAAGAA
TACAGAGGACTTCCATACAACCATGCCAATTGTGAGATCCAGGCAATGAAAACGCCCCTG
AGGCCTTTCAGTGACGATATCAACCACAACCCAGTCACAAAGGCTAACGCGAAGCCATTA
GATGTGTTCGAGTATAATCGGTTGAGCTTCCAGTACGACAACCTCATCTTCCATGGATAC
AGTATTCCGGAACTTGATCGCGTGCTTGAAGAAAGAAAGGAGGAGGACAGAATATTTGCT
GCCTTCCTTCTCAGTGGAATCAAGCGTAGTGCTGATGTAGTGTTCGACATATGCCAGCCA
GAACACGAATGTGTGTTCGCAGGGACTTTTGCGATTTTGGGAGGGGAGCTAGAAATGCCC
TGGTCCTTCGACAGACTGTTCCGCTATGATATCACCAAGGTGATGAAGCAGCTACACCTG
AGGCATGACTCTGACTTTACCTTCAGGGTGAAGATTGTCGGCACCGACGACCACGAGCTT
CCTTCAGACAGTGTCAAAGCACCAACTATTGAATTTGAACCGGGCG

INTRON 1G-3/1H (SEQ ID NO:123)

GTGAGTACGACAGGCATTTCTAGTAAAAACCTACTTTTGGTAAAAGGTTCGAGAAATCAC
TTGAAGCAACAACATGATTTTGTAACGCCTATTACACGTGAACATGTCACACCCGGTGAT
GCCGTTTAATGGACATGCCTCTGTTAATGAAAGGGGTAAGTACATGTGTATGGGGATGGG
ATGGGAGCCACCTGTCCCAATTTCATAGGTCCCTAGGATCCCAGTTGCGTAGGAATCCCC
TGATTAATGCCTTGTGAATTCCTCCTGGAATTGTCCTGGCCCAAATTTTTACAAACCCGC
CCCGATATACCTTGGAAATAATTGGGCCTAAGGGTGGGGCTTTTAAGGACCAAGAACCCA
ACCTAAACCCCAACCCATTTTTTCCCACCCATTCCAGGTTTTGTTTTACCAAATAAAAAG
GTTTCCACTTTGAGGAAACCCTTTAAGGGTTCTTTTCAGGGCTTTTTTTCTTTTCTGGGA
ATTCCAATTCCGGGGGAACAAAATACATATATTTCACAGACCTTTGGTCAAATTTATATA
ATTTCCGACTTCATGTCATAGGTTTGTCTTTCTTCCTACACAG

DOMAIN 1H

TGCACAGAGGCGGAAACCACGAAGATGAACACCATGATGACAGACTCGCAGATGTCCTGA
TCAGGAAAGAAGTTGACTTCCTCTCCCTGCAAGAGGCCAACGCAATTAAGGATGCACTGT
ACAAGCTCCAGAATGACGACAGTAAAGGGGGCTTTGAGGCCATAGCTGGCTATCACGGGT
ATCCTAATATGTGTCCAGAAAGAGGTACCGACAAGTATCCCTGCTGTGTCCACGGAATGC
CCGTGTTCCCCCACTGGCACCGCCTGCATACCATTCAGATGGAGAGAGCTCTGAAAAACC
ATGGCTCTCCAATGGGCATTCCTTACTGGGATTGGACAAAGAAGATGTCGAGTCTTCCAT
CTTTCTTTGGAGATTCCAGCAACAACAACCCTTTCTACAAATATTACATCCGGGGCGTGC
AGCACGAAACAACCAGGGACATTAATCAGAGACTCTTTAATCAAACCAAGTTTGGTGAAT
TTGATTACCTATATTACCTAACTCTGCAAGTCCTGGAGGAAAACTCGTACTGTGACTTTG
AAGTTCAGTATGAGATCCTCCATAACGCCGTCCACTCCTGGCTTGGAGGAACTGGAAAGT
ATTCCATGTCTACCCTGGAGCATTCGGCCTTTGACCCTGTCTTCATGATTCACCACTCGA
GTTTGGATAGAATCTGGATCCTTTGGCAGAAGTTGCAAAAGATAAGAATGAAGCCTTACT
ACGCATTGGATTGTGCTGGCGACAGACTTATGAAAGACCCCCTGCATCCCTTCAACTACG
AAACCGTTAATGAAGATGAATTCACCCGCATCAACTCTTTCCCAAGCATACTGTTTGACC
ACTACAGGTTCAACTATGAATACGATAACATGAGAATCAGGGGTCAGGACATACATGAAC
TTGAAGAGGTAATTCAGGAATTAAGAAACAAAGATCGCATATTTGCTGGTTTTGTTTTGT
CGGGCTTACGGATATCAGCTACAGTGAAAGTATTCATTCATTCGAAAAACGATACAAGTC
ACGAAGAATATGCAGGAGAATTTGCAGTTTTGGGAGGTGAGAAGGAGATGCCGTGGGCAT
ATGAAAGAATGCTGAAATTGGACATCTCCGATGCTGTACACAAGCTTCACGTGAAAGATG
AAGACATCCGTTTTAGAGTGGTTGTTACTGCCTACAACGGTGACGTTGTTACCACCAGGC
TGTCTCAGCCATTCATCGTCCACCGTCCAGCCCATGTGGCTCACGACATCTTGGTAATCC
CAGTAGGTGCGGGCCATGACCTTCCGCCTAAAGTCGTAGTAAAGAGCGGCACCAAAGTCG
AGTTTACACCAATAGATTCGTCGGTGAACAAAGCAATGGTGGAGCTGGGCAGCTATACTG
CTATGGCTAAATGCATCGTTCCCCCTTTCTCTTACCACGGCTTGAACTGGACAAAGTCT
ACAGCGTCGATCACGGAGACTACTACATTGCTGCAGGTACCCACGCGTTGTGTGAGCAGA
ACCTCAGGCTCCACATCCACGTGGAACACGAGTAG

3´UTR

TTCACAG

INTRON 3´UTR (SEQ ID NO:124)

GTGAGGAGAAGGCCCCAGGCTAGCAGGGCAATGGATGAAGGAAATAGGGGCAAAGGGAAT
AGCAGTTACACCATCGACATTTCCAACCTCCTCAGAAACTAATATATAGCCTTAATACAA
CCAGCCAAGACTCAACGGGCAGCCGGGGTGGGGGGATTTGGTGGTCGCTGTTTCAGACCA
GGGTGCAAAATATCAGTGCGCAAATCAACATGTTGCGTGTCAGACACTGACACAGCAGTC
ATTGAACCTGCAGACCCATAACAGGAAAATGGGGCAGATACGATCAAAGACAGTGTAAAA
TAGGGATAAGTAGGCATATGCAACCACCTGATGGAAATGAAAAGGGGTAAGTTTAAACCC
CGGCTACCAAAGGTCCAATGGTTCCTTAACCCAGCTTACGCTATCCCTCTAATTTCAGTA
TTGAGCTGATTTCTGTCGAGTTCATGTAAACTGTATACTTTCTGTATTATTACAG

3´UTR

GTTGCTATGCCGACTGCGCTATATTGGTGAACGAGACGATGAGGACATCTCTGAAAGAGT
TCGCCAAGTGATGTGTAGGTCACGGAAGTATTGTTGAGCTAACAATATGATGATTTCAAA
ATGACTTGGCGCTCTAGGACAAAGACATAATTCATCAGCACCCTGTGCACCAACTCTTTG
TTTGCTGCAAACGTCTGACAAGCGACACGTCAATCAACAAGCTGTTCAAACTCAAGTGGA
TGTAACTAGAATCGTTGGGCCATCGTTCACAAAGTATTGACAGATGTCACACATGATGGC
GAGAAACACTTTAGAACTTTTAATGACCTAGAGTGACTTGTAAATATGTAAATATATTCT

TCAAAGACTCAGCTGAACTATTGTTGGATAACACATCAATTCCCTCAACAAAATGCTTTA
TCTTCACATGGATGTATGTAATGTGGCCGGCAATAAAGTATATATATGTAT

Figure 5

Primary structure of the HtH1 protein

SIGNAL PEPTIDE

LVQFLLVALVVGAGA

DOMAIN A

DNVVRKDVSHLTVDEVQALHGALHDVTASTGPLSFEDITSYHAAPASCDYKGRKIACCVHGMPSFP
FWHRAYVVQAERALLSKRKTVGMPYWDWTQTLTHLPSLVTEPIYIDSKGGKAQTNYWYRGEIAFIN
KKTARAVDDRLFEKVEPGHYTHLMETVLDALEQDEFCKFEIQFELAHNAIHYLVGGKFEYSMSNLE
YTSYDPIFFLHHSNVDRLFAIWQRLQELRGKNPNAMDCAHELAHQQLQPFNRDSNPVQLTKDHSTP
ADLFDYKQLGYSYDSLNLNGMTPEQLKTELDERHSKERAFASFRLSGFGGSANVVVYACVPDDDPR
SDDYCEKAGDFFILGGQSEMPWRFYRPFFYDVTEAVHHLGVPLSGHYYVKTELFSVNGTALSPDLL
PQPTVAYRPGK

DOMAIN B

GHLDPPVHHRHDDDLIVRKNIDHLTREEEYELRMALERFQADTSVDGYQATVEYHGLPARCPRPDA
KVRFACCMHGMASFPHWHRLFVTQVEDALVRRGSPIGVPYWDWTKPMTHLPDLASNETYVDPYGHT
HHNPFFNANISFEEGHHHTSRMIDSKLFAPVAFGEHSHLFDGILYAFEQEDFCDFEIQFELVHNSI
HAWIGGSEDYSMATLHYTAFDPIFYLHHSNVDRLWAIWQALQIRRHKPYQAHCAQSVEQLPMKPFA
FPSPLNNNEKTHSHSVPTDIYDYEEVLHYSYDDLTFGGMNLEEIEEAIHLRQQHERVFAGFLLAGI
GTSALVDIFINKPGNQPLKAGDIAILGGAKEMPWAFDRLYKVEITDSLKTLSLDVDGDYEVTFKIH
DMHGNALDTDLIPHAAVVSEPAH

DOMAIN C

PTFEDEKHSLRIRKNVDSLTPEETNELRKALELLENDHTAGGFNQLGAFHGEPKWCPNPEAEHKVA
CCVHGMAVFPHWHRLLALQAENALRKHGYSGALPYWDWTRPLSQLPDLVSHEQYTDPSDHHVKHNP
WFNGHIDTVNQDTTRSVREDLYQQPEFGHFTDIAQQVLLALEQDDFCSFEVQYEISHNFIHALVGG
TDAYGMASLRYTAYDPIFFLHHSNTDRIWAIWQSLQKYRGKPYNTANCAIESMRRPLQPFGLSSAI
NPDRITREHAIPFDVFNYRDNLHYVYDTLEFNGLSISQLDRELEKIKSHERVFAGFLLSGIKKSAL
VKFEVCTPPDNCHKAGEFYLLGDENEMAWAYDRLFKYDITQVLEANHLHFYDHLFIRYEVFDLKGV
SLGTDLFHTANVVHDSGT

DOMAIN D

GTRDRDNYVEEVTGASHIRKNLNDLNTGEMESLRAAFLHIQDDGTYESIAQYHGKPGKCQLNDHNI
ACCVHGMPTFPQWHRLYVVQVENALLNRGSGVAVPYWEWTAPIDHLPHFIDDATYFNSRQQRYDPN
PFFRGKVTFENAVTTRDPQAGLFNSDYMYENVLLALEQENYCDFEIQFELVHNALHSMLGGKGQYS
MSSLDYSAFDPVFFLHHANTDRLWAIWQELQRFRELPYEEANCAINLMHQPLKPFSDPHENHDNVT
LKYSKPQDGFDYQNHFGYKYDNLEFHHLSIPSLDATLKQRRNHDRVFAGFLLHNIGTSADITIYIC
LPDGRRGNDCSHEAGTFYILGGETEMPFIFDRLYKFEITKPLQQLGVKLHGGVFELELEIKAYNGS
YLDPHTFDPTIIFEPGT

DOMAIN E

DTHILDHDHEEEILVRKNIIDLSPRERVSLVKALQRMKNDRSADGYQAIASFHALPPLCPNPSAAH
RYACCVHGMATFPQWHRLYTVQVQDALRRHGSLVGIPYWDWTKPVNELPELLSSATFYHPIRNINI
SNPFLGADIEFEGPGVHTERHINTERLFHSGDHDGYHNWFFETVLFALEQEDYCDFEIQFEIAHNG

IHTWIGGSAVYGMGHLHYASYDPIFYIHHSQTDRIWAIWQELQKYRGLSGSEANCAIEHMRTPLKP
FSFGPPYNLNSHTQEYSKPEDTFDYKKFGYRYDSLELEGRSISRIDELIQQRQEKDRTFAGFLLKG
FGTSASVSLQVCRVDHTCKDAGYFTILGGSAEMPWAFDRLYKYDITKTLHDMNLRHEDTFSIDVTI
TSYNGTVLSGDLIQTPSIIFVPGR

DOMAIN F

HKLNSRKHTPNRVRHELSSLSSRDIASLKAALTSLQHDNGTDGYQAIAAFHGVPAQCHEPSGREIA
CCIHGMATFPHWHRLYTLQLEQALRRHGSSVAVPYWDWTKPITELPHILTDGEYYDVWQNAVLANP
FARGYVKIKDAFTVRNVQESLFKMSSFGKHSLLFDQALLALEQTDYCDFEVQFEVMHNTIHYLVGG
RQTYAFSSLEYSSYDPIFFIHHSFVDKIWAVWQELQSRRHLQFRTADCAVGLMGQAMRPFNKDFNH
NSFTKKHAVPNTVFDYEDLGYNYDNLEISGLNLNEIEALIAKRKSHARVFAGFLLFGLGTSADIHL
EICKTSENCHDAGVIFILGGSAEMHWAYNRLYKYDITEALQEFDINPEDVFHADEPFFLRLSVVAV
NGTVIPSSHLHQPTIIYEPGE

DOMAIN G

DHHDDHQSGSIAGSGVRKDVNTLTKAETDNLREALWGVMADHGPNGFQAIAAFHGKPALCPMPDGH
NYSCCTHGMATFPHWHRLYTKQMEDAMRAHGSHVGLPYWDWTAAFTHLPTLVTDTDNNPFQHGHID
YLNVSTTRSPRDMLFNDPEHGSESFFYRQVLLALEQTDFCKFEVQFEITHNAIHSWTGGHSPYGMS
TLDFTAYDPLFWLHHSNTDRIWAVWQALQEYRGLPYNHANCEIQAMKTPLRPFSDDINHNPVTKAN
AKPLDVFEYNRLSFQYDNLIFHGYSIPELDRVLEERKEEDRIFAAFLLSGIKRSADVVFDICQPEH
ECVFAGTFAILGGELEMPWSFDRLFRYDITKVMKQLHLRHDSDFTFRVKIVGTDDHELPSDSVKAP
TIEFEPG

DOMAIN H

VHRGGNHEDEHHDDRLADVLIRKEVDFLSLQEANAIKDALYKLQNDDSKGGFEAIAGYHGYPNMCP
ERGTDKYPCCVHGMPVFPHWHRLHTIQMERALKNHGSPMGIPYWDWTKKMSSLPSFFGDSSNNNPF
YKYYIRGVQHETTRDINQRLFNQTKFGEFDYLYYLTLQVLEENSYCDFEVQYEILHNAVHSWLGGT
GKYSMSTLEHSAFDPVFMIHHSSLDRIWILWQKLQKIRMKPYYALDCAGDRLMKDPLHPFNYETVN
EDEFTRINSFPSILFDHYRFNYEYDNMRIRGQDIHELEEVIQELRNKDRIFAGFVLSGLRISATVK
VFIHSKNDTSHEEYAGEFAVLGGEKEMPWAYERMLKLDISDAVHKLHVKDEDIRFRVVVTAYNGDV
VTTRLSQPFIVHRPAHVAHDILVIPVGAGHDLPPKVVVKSGTKVEFTPIDSSVNKAMVELGSYTAM
AKCIVPPFSYHGFELDKVYSVDHGDYYIAAGTHALCEQNLRLHIHVEHE

Figure 6

Genomic sequence of the HtH2 gene

DOMAIN 2A-1 (1st part of domain a)
[domain a, parts 1-4: SEQ ID NO:156]

GGTCTTCCGTACTGGGACTGGACGCAGCATCTGACTCAACTCCCAGATCTGGTGTCAGACCCCTTG
TTTGTCGACCCGGAAGGAGGAAAG

INTRON 2A-1/2A-2 (SEQ ID NO:125)

GTAAGGGATCTCAGATCCGTCAGAGTGAGTGAGTGAGTGAGTGAGTGCCCAGCAACTGAAGCTAGG
CCGCCCTACTGGGGATCACAGGGAATGTATGTCAATGGTTGAAGAAAGGAGCAGTGGGTTACAACG
CCGCGTTCAAAGTCATGGCAGTTTCATAGCGCATTGTGCGCGCGTGTGTATCTGTGTGCGCGCGTG
TGTGCTTGCGTGCGTGTGAGTGAGTCCGCTTGTGCATTTGTACTAGCACAGACTAATGCTGGTTCT
AGAGAGCCTACTGATAAATGTTTACATTAAGATCTTTACAGTATACTGAGATTCGAGCCCAGACCA
GCGGAACACCAGGCAGGGTAACAACAAATAACGCCTTTCCACACAACCGACGCAGCCTAAAGTGGC
TCTGATAGGCTGATACCGGTGTATTCTTAGAACTTGTAATTTGTGCTTTGCCATAATACATGTACT
TCAGTTAACTGTAATACAGCATAAGACTGGACCGGTGTTTACGACGCAATGAGCAATAATTACTCT
ACGAAAAGATTTGGTTAGACATATTCAATAATTGTAACATTCATTAACAATGAACACCACGTGCAC
TCTCGTTTGTGTCAACGTATTCATAATCATTCTCATGCATCTGTTAGCTCAGATATTTTGATGTTT
CAAGAGATTTGTACGAACGTATGGGCTGGTGCCCCATGAAATTACATACAATGAATTCAGGTGAAA
TACCTGGCGAGACAATAAGATCTTACTAGTGCTGCCACTTCAGTATGGTGTCCCCGATGGTGTCTG
GTGTATGGGTGTGTTTGGCGTCAGTTGTTACTGGAAAAGTCAGCTCTAATTATGTCTTTATGTGGT
TAAAGACCCCATAACCTAGATGTCTGGGTTTAACTTAACATGATAGTAACAGTCGGCTGTATAGCC
TGACGCTTAAACGTTAGATGAATAAGGACTATATTGTGTTGTATAACATTTCTATAACCTCCTTTC
TATATCATTTAG

DOMAIN 2A-2 (2nd part of domain a)

GCCCATGACAACGCATGGTATCGTGGAAACATCAAGTTTGAGAATAAGAAGACTGCAAGAGCTGTT
GACGATCGCCTTTTCGAGAAGGTTGGACCAGGAGAGAATACCCGACTCTTTGAAGGAATTCTCGAT
GCTCTTGAACAGGATGAATTCTGCAACTTCGAGATCCAGTTTGAGTTGGCTCACAACGCTATCCAC
TACCTGGTTGGCGGCCGTCACAC

INTRON 2A-2/2A-3 (SEQ ID NO:126)

GTGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATG
GTCACGAGTCACGTTCTCTGATGGTCACGAGTCACATTCTCTGATGGTCACGAGTCACATTCTCTG
TTGAGTGAAGTCTCAGTACCATTTATTTCTCTTACCTTCTTCTAACCAGGGGTTTCAGCGTGGATC
GTCTGAGAAGTTAGCGCAAATCTATATTGAAGTCATTTTTCTATCATATAACCATCGTTATATCCA
CGTGCGAAAGTGTTCATTAATTATTTTTATTTTCATTTATGAAGGTCTAAAAGAAAATATGTATTG
TTGGAAACTATATTCGAAGGTGAAGGCAACACGAGTGTATTAATATTCTCAATATCAATGTACGCT
CTGTCAGCACCTGTTTCACCAGGAACTACACCTTTAGCGTACCAAAATATCAGCTGATGATTTCGA
AGCGGACTATACCCTCACCACTTGTTTTGTGTGTGTATTTATGTGTGCATGTGTGTGCGTGCGTGC
GTGTGTGTGTGTGTCCTACGTATGTTGATATTTTGTTCTGACTGTATATGTTCGTGCTTACCATTG
AAG

DOMAIN 2A-3 (3rd part of domain a)

GTACTCCATGTCTCATCTCGAGTACACCTCCTACGACCCCCTCTTCTTCCTCCATCACTCCAACAC
CGACCGCATCTTCGCCATCTGGCAACGTCTTCAGGTACTCAGAGGAAAGGACCCCAACACCGCCGA
CTGCGCACACAACCTCATCCATGAGCCCATGGAACCGTTCCGTCGGGACTCGAACCCTCTTGACCT
CACCAGGGAAAACTCCAAACCAATTGACAGCTTTGATTATGCCCACCTTGGCTACCA

INTRON 2A-3/2A-4 (SEQ ID NO:127)

GTATGTATGATTCTAATAATGAATGTTTTTACCTCCGGTTTAAACAATATTTTAGTATTACGAAAG
GAGAAGTACCTCGAGAGGTCTAGGTCTCAGATGTTTAGAAACCCATGAAGACAGGTATGCTTCTGA
AAAACAAAGTAACATCATGAGGCTAAAGTTCAGATTCAAACCATCGTAGTTCGAATCCAGCATGCA
AAGGGCCCTAACCCTGTAGATGGCGCTGCTTGAAACAGAGTAGTCTGTTCAGGGTCAGTACTGTCC
CCACAAACATCATAGTCAGGGTCAGTACTGTCCCCACAAACATCATAGTCAGGGTCAGTACTGTCC
CCACAAACATCACAGTCAGGGTTAATTTTGGATTCGGTTTCGAATGCGAAGAAGACAGTCACGCCC
TGACACTGGACCGAGGTTGCCGAGAAAGCTCGTGATATTGCTGGAATACTGCCCAGTAAAACCATC
ATTTATTTTAGGCTATTTATTACGAAAAATAATAATATGTATAGAAATGCATATGATCGCTGTTTG
AATGTAAAATTTAGAATGGGTTTGGGAGTGTTCACTATTTTTTCATCAAAATTTCATGTATTTTAA
CCGATCGACGCTGAAGACAAACTACCGTTAATCAGGCAGTTCATTCATATCTGATAGGGAATATTG
GTTGTTAACCAACGCTACATTGTGTCCAG

DOMAIN 2A-4 (4th part of domain a)

GTATGATGACTTGACCCTGAACGGTATGACCCCAGAGGAATTGAACTCATATCTGCATGAACGGTC
AGGCAAGGAGGGGGTGTTCGCAAGCTTCCGACTCTCAGGTTTTGGCGGCTCTGCTAACGTTGTTGT
CTACGCATGCCGTCCTGCCCACGATGAAATGGCTGTCGATCAGTGCGACAAAGCCGGCGACTTCTT
TGTGTTGGGCGGACCCACCGAGATGCCCTGGAGGTTTACAGAGCATTCCACTTCGACGTCACCGA
CAGCATCGACAACATCGACAAGGACCGCCACGGCCACTATTATGTAAAGGCGGAATTATTCAGTGT
AAATGGAAGTGCGCTACCGAATGATCTCCTGCCTCAACCCACCATCTCACACAGGCCAGCCCGCGG
ACACGTTGATG

INTRON 2A-4/2B (SEQ ID NO:128)

GTAAATGGCCATTGTATACATGCATTCATTTGGACTTTGAGTGAGTGAGTGGATGCGTATTCAGTA
AGTGAGAGTGTGAGTGGGTATTAGGTCTGTGAGTGGGTTGGTGAGTGGATGGGTGAGTAAGAGTGG
GTTGGTGAGAAAGTGAGTGAGTCACTTGGTGGGTGCGTTAGTGGAAGCGTGATTGAGTGGATGGGA
GGTAGGTGAGTGAGTGAATTGGTGGGGGGGTGAGTGAGGTTAACGCTGTTCTGCTGTTCAATCACA
CCACATGTTGCCAGCTTACTGTGCAGGACGAATCCAGGGTTGTGTTAAATTTTATATGTTTATATA
TAACGATGGACGTGTCTGGATGTGGCGAATGTGTCAAGAGAATTATGCGGCTTTGTGCTGCTCCGC
GTATTTATTGCACGCGCGTTGGTACGCGGTTGATAAAGTAGTTCAAAACATTTCCCAGCCATCTTT
GTCTGTTGTGAAAACCTACTCCAGGACCATCCATTTCAATATGTGTCTGCGTTCATGGAGTTATAC
ATGTTAAACTGTAGAGCGCAGATGAGCACACTTGAGCATTTCTTCAGTAAATCAGAATGTGTATAT
TTCAAAATTTACCAAATGCAATATCATCAAGCAAATTATGCAGCTCTATAGTAACATCGGAGTCAA
TGGTCCAGTGTGCCCTCGGCTGCCATTCCGACCTCCCTGGCCAGAATACACCCCGGTCAGGATCAG
TTATCCGTCAGAAGGCACGGTGCGGAATGAAAACATAAACACATAGTCGCTTAGTAGTATGCTGAT
TTAGGCACGCAAAATCCGAATGTGAATTACTGTGAATTGCATTACCTGTTACAG

DOMAIN 2B

AGGCCCCAGCTCCCTCCTCGGATGCTCACCTCGCCGTCAGGAAGGATATCAACCATCTGACACGCG
AGGAGGTGTACGAGCTGCGCAGAGCTATGGAGAGATTCCAGGCCGACACATCCGTTGATGGGTACC
AGGCTACGGTTGAGTATCACGGCTTACCTGCTCGATGTCCATTCCCCGAGGCCACAAATAGGTTCG
CCTGTTGCATCCACGGCATGGCGACATTCCCTCATTGGCACAGACTGTTCGTTACCCAGGTGGAAG
ATGCACTGATCAGGCGAGGATCCCCTATAGGGGTCCCCTACTGGGACTGGACTCAGCCTATGGCAC
ATCTCCCAGGACTTGCAGACAACGCCACCTATAGAGATCCCATCAGCGGAGACAGCAGACACAACC
CGTTCCACGATGTTGAAGTTGCCTTTGAAAATGGGCGTACAGAACGTCACCCAGATAGTAGATTGT

TTGAACAACCTCTATTTGGCAAACATACGCGTCTCTTCGACAGTATAGTCTATGCTTTTGAGCAGG
AGGACTTCTGCGATTTTGAAGTTCAATTTGAGATGACCCATAATAATATTCACGCCTGGATTGGTG
GCGGCGGGAAGTATTCCATGTCTTCTCTACACTACACAGCCTTCGACCCTATCTCCTACCTTCATC
ACTCCAACACTGACCGTCTCTGGGCAATTTGGCAAGCGTTGCAGATACGAAGAAACAAACCGTATA
AGGCTCATTGTGCTTGGTCTGAGGAACGCCAGCCTCTCAAACCTTTCGCCTTCAGTTCCCCACTGA
ACAACAACGAAAAAACCTACGAAAACTCGGTGCCCACCAACGTTTACGACTACGAAGGAGTCCTTG
GCTATACTTATGATGACCTCAACTTCGGGGGCATGGACCTGGGTCAGCTTGAGGAATACATCCAGA
GGCAGAGACAGAGAGACAGGACCTTTGCTGGCTTCTTTCTGTCACATATTGGTACATCAGCGAATG
TTGAAATCATTATAGACCATGGGACTCTTCATACCTCCGTGGGCACGTTTGCTGTTCTTGGCGGAG
AGAAGGAGATGAAATGGGGATTTGACCGTTTGTACAAATATGAGATTACAGATGAACTGAGGCAAC
TTAATCTCCGTGCTGATGATGGTTTCAGCATCTCTGTTAAAGTAACTGATGTTGATGGCAGTGAGC
TGTCCTCTGAACTCATCCCATCTGCTGCTATCATCTTCGAACGAAGCCATA

INTRON 2B/2C (SEQ ID NO:129)

GTAAGTAGCTACCTGTTTATTCAATTTTTTCGCTTTGCCAATCAATTCATTCAGCTTGAAATTCAA
TAATTGTGTTTTGCATGGCTGAAAACCAATTTGAACTCTTTTCTTTTCTCAGGTCGAACTCAAATA
AATAATCACTAATTGTTATGCACGCGGGTAGGGCATACATACTATATCCACATCGGTCATCTCAAA
ATGCAAACAAATTGTCTTATTTCCGTTGGGACAAGCAAACCCCCTTTCCTGTAATCTTGCCTTTGG
CATCCACTGGAATTAATGTTGACTGGTAATTGATACTGGCTCTCTTCTTGCATAGAGTTAATATCT
ATAGTTTGTAAATCTTTATGATTTTGCTATTTATATTTCGACAGCATGCTATAGACACCCTAGACT
ATTGTATAGCCACTTGTATTGTTTTTCCATTTATTATTTATAACAGAACATGGCTTGTAATTTTTA
TTTACCTTCCAG

DOMAIN 2C

TTGACCATCAGGACCCTCATCAGGACACAATCATCAGGAAAAATGTTGATAATCTTACACCCGAGG
AAATTAATTCTCTGAGGAGGGCAATGGCAGACCTTCAATCAGACAAAACCGCCGGTGGATTCCAGC
AAATTGCTGCTTTTCACGGGGAACCCAAATGGTGCCCAAGTCCCGATGCTGAGAAGAAGTTCTCCT
GCTGTGTCCATGGAATGGCTGTCTTCCCTCACTGGCACAGACTCCTGACCGTGCAAGGCGAGAATG
CCCTGAGAAAGCATGGATGTCTCGGAGCTCTCCCCTACTGGGACTGGACTCGGCCCCTGTCTCACC
TACCTGATTTGGTAAGTCAGCAGAACTACACCGATGCCATATCCACCGTGGAAGCCCGAAACCCCT
GGTACAGCGGCCATATTGATACAGTTGGTGTTGACACAACAAGAAGCGTCCGTCAAGAACTGTATG
AAGCTCCCGGATTTGGTCATTATACTGGGGTCGCTAAGCAAGTGCTTCTGGCTTTGGAGCAGGATG
ACTTCTGTGATTTTGAAGTCCAGTTTGAGATAGCTCACAATTTCATCCACGCTCTTGTCGGCGGAA
GCGAGCCATATGGTATGGCGTCACTCCGTTACACTACTTATGATCCAATTTTCTACCTCCATCATT
CTAACACTGACAGACTCTGGGCTATATGGCAGGCTCTACAAAAGTACAGGGGCAAACCTTACAATT
CCGCCAACTGTGCCATTGCTTCTATGAGAAAACCCCTACAGCCCTTTGGTCTGACTGATGAGATCA
ACCCGGATGATGAGACAAGACAGCATGCTGTTCCTTTCAGTGTCTTTGATTACAAGAACAACTTCA
ATTATGAATATGACACCCTTGACTTCAACGGACTATCAATCTCCCAGCTGGACCGTGAACTGTCAC
GGAGAAAGTCTCATGACAGAGTATTTGCCGGATTTTTGCTGCATGGTATTCAGCAGTCTGCACTAG
TTAAATTCTTTGTCTGCAAATCAGATGATGACTGTGACCACTATGCTGGTGAATTCTACATCCTTG
GTGATGAAGCTGAAATGCCATGGGGCTATGATCGTCTTTACAAATATGAGATCACTGAGCAGCTCA
ATGCCCTGGATCTACACATCGGAGATAGATTCTTCATCAGATACGAAGCGTTTGATCTTCATGGTA
CAAGTCTTGGAAGCAACATCTTCCCCAAACCTTCTGTCATACATGACGAAGGGGCAG

INTRON 2C/2D (SEQ ID NO:130)

GTGAGAACATTGATAATAGTTCAAATGAAGTATATCCGATTCAAGCTGTCGATACAAGATGAGATA
CATAATCACAATGTTTGTATTAGATATCTCTCTTAATTTAATGCCGCTTTTATCAATATTCGAGCA
ATCCTTCAGCAACATACACCAGCAAATGTTTCATCAACAGACTATATTATTTAATATTTTAAAAAT
CCTTCTCTGTTGTTATAAATACTTAAAGTATCGAATTCCTTGAATGCGTCTTCTCTGCAGCATATA
GTTAAGTTGTTGTGTTTCTCTGTCAG

DOMAIN 2D

GTCACCATCAGGCTGACGAGTACGACGAAGTTGTAACTGCTGCAAGCCACATCAGAAAGAATTTAA
AAGATCTGTCAAAGGGAGAAGTAGAGAGCCTAAGGTCTGCCTTCCTGCAACTTCAGAACGACGGAG
TCTATGAGAATATTGCCAAATTCCACGGCAAGCCTGGGTTGTGTGATGATAACGGTCGCAAGGTTG
CCTGTTGTGTCCATGGAATGCCCACCTTCCCCCAGTGGCACAGACTCTATGTCCTCCAGGTGGAGA
ATGCTTTGCTGGAGAGAGGATCTGCCGTCTCTGTGCCATACTGGGACTGGACTGAAACATTTACAG
AGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCAATTCCCGTCAACAAACGTTTGACCCTAATC
CTTTCTTCAGAGGTAAAATCAGTTTTGAGAATGCTGTTACAACACGTGATCCCCAGCCTGAGCTGT
ACGTTAACAGGTACTACTACCAAAACGTCATGTTGGCTTTTGAACAGGACAACTACTGCGACTTCG
AGATACAGTTTGAGATGGTTCACAATGTTCTCCATGCTTGGCTTGGTGGAAGAGCTACTTATTCTA
TTTCTTCTCTTGATTATTCTGCATTCGACCCTGTGTTTTTCCTTCACCATGCGAACACAGATAGAT
TGTGGGCCATCTGGCAGGAGCTGCAGAGGTACAGGAAGAAGCCATACAATGAAGCGGATTGTGCCA
TTAACCTAATGCGCAAACCTCTACATCCCTTCGACAACAGTGATCTCAATCATGATCCTGTAACCT
TTAAATACTCAAAACCCACTGATGGCTTTGACTACCAGAACAACTTTGGATACAAGTATGACAACC
TTGAGTTCAATCATTTCAGTATTCCCAGGCTTGAAGAAATCATTCGTATTAGACAACGTCAAGATC
GTGTGTTTGCAGGATTCCTCCTTCACAACATTGGGACATCCGCAACTGTTGAGATATTCGTCTGTG
TCCCTACCACCAGCGGTGAGCAAAACTGTGAAAACAAAGCCGGAACATTTGCCGTACTCGGAGGAG
AAACAGAGATGGCGTTCATTTTGACAGACTCTACAGGTTTGACATCAGTGAAACACTGAGGGACC
TCGGCATACAGCTGGACAGCCATGACTTTGACCTCAGCATCAAGATTCAAGGAGTAAATGGATCCT
ACCTTGATCCACACATCCTGCCAGAGCCATCCTTGATTTTTGTGCCTGGTTCAA

INTRON 2D/2E (SEQ ID NO:131)

GTAAGAAAGTTTCACTGTCTAAATCTTTTTTTATGATAGAGGGTAGAGAAGTGGAGACAATGTGAC
AATATATTGAATAAAGTTGTTTAAAATTTATAACTCTCATAAGTTCATATTATGCTGAAGCTGTAG
CCATCTATAACTGTGTAACATGAAATGTTAAGACATTAACCTAAATACTTCAGCTGATAACAAAAC
AATGTTAATACATACGTCAATGTAACATTTCTTATCTTTAGGTTATAGCATAAACACTTCAGAGA
TACAGTGACGAAAACCTCTATTTAAATATTTCAG

DOMAIN 2E

GTTCTTTCCTGCGTCCTGATGGGCATTCAGATGACATCCTTGTGAGAAAAGAAGTGAACAGCCTGA
CAACCAGGGAGACTGCATCTCTGATCCATGCTCTGAAAAGTATGCAGGAAGACCATTCACCTGATG
GGTTCCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGCCCTTCACCATCTGCAACTCACC
GTTATGCTTGCTGTGTCCACGGCATGGCTACATTTCCCCAGTGGCACAGACTGTACACTGTACAGT
TCCAGGATGCACTGAGGAGACATGGAGCTGCAGTAGGTGTACCGTATTGGGATTGGCTGCGACCGC
AGTCTCACCTACCAGAGCTTGTCACCATGGAGACATACCATGATATTTGGAGTAACAGAGATTTCC
CCAATCCTTTCTACCAAGCCAATATTGAGTTTGAAGGAGAAAACATTACAACAGAGAGAGAAGTCA
TTGCAGACAAACTTTTTGTCAAAGGTGGACACGTTTTTGATAACTGGTTCTTCAAACAAGCCATCC
TAGCGCTTGAGCAGGAAAACTACTGTGACTTTGAGATTCAGTTTGAAATTCTTCACAACGGCGTTC
ACACGTGGGTCGGAGGCAGTCGTACCCACTCTATCGGACATCTCCATTACGCATCCTACGACCCTC
TTTTCTACCTCCACCATTCCCAGACAGACCGTATTTGGGCAATCTGGCAAGAACTCCAGGAACAGA
GAGGGCTCTCAGGTGATGAGGCTCACTGTGCTCTCGAGCAAATGAGAGAACCATTGAAGCCTTTCA
GCTTCGGCGCTCCTTATAACTTGAATCAGCTAACACAGGATTTCTCCCGACCCGAGGACACCTTCG
ACTACAGGAAGTTTGGTTATGAATATGACAATTTAGAATTCCTAGGAATGTCAGTTGCTGAACTGG
ATCAATACATTATTGAACATCAAGAAAATGATAGAGTATTCGCTGGGTTCCTGTTGAGTGGATTCG
GAGGTTCCGCATCAGTTAATTTCCAGGTTTGTAGAGCTGATTCCACATGTCAGGATGCTGGGTACT
TCACCGTTCTTGGTGGCAGTGCTGAGATGGCGTGGGCATTTGACAGGCTATACAAATATGACATTA
CTGAAACTCTGGAGAAAATGCACCTTCGATATGATGATGACTTCACAATCTCTGTCAGTCTGACCG
CCAACAACGGAACTGTCCTGAGCAGCAGTCTAATCCCAACACCGAGTGTCATATTCCAGCGGGGAC
ATC

INTRON 2E/2F-1 (SEQ ID NO:132)

GTAAGTAGTAAACTGCTCAGATTGTTTTCATAATTACTCCACTATTAAGTAAAAAGTACTAGTAAT
TCAATAGTACTGTTCACAGAGAAATGTAACACAATAGACCACAGAGTCCATTTGTTAAACGCCTTT
GGCTTGGTAAGTCTGAGATTTTGGTGACTGATGGAAAGCTAAAATATATTTTGACAG

DOMAIN 2F-1 (1st part of domain f)

GTGACATAAATACCAAGAGCATGTCAGCGAACCGTGTTCGCCGTGAGCTGAGCGATCTGTCTGCGA
GGGACCCGTCTAGTCTCAAGTCTGCTCTGCGAGACCTACAGGAGGATGATGGCCCCAACGGATACC
AGGCTCTTGCAGCCTTCCATGGGCTACCAGCAGGCTGCCATGATAGCCAGGGAAATGAG

INTRON 2F-1/2F-2 (SEQ ID NO:133)

GTATATTTAAGTATTTTATCTTACGCATGACCCTGACCCTATTTATTTTTTTTAATCCTCGGATT
TGTTTAATCCTGTTACCAGCGAAGGTCCGGGTTAGAATTGATCTTCAGTCAACTATTCTTGTCGTA
GGACTAACGAGTTGTCTGGCTTGCTTACTCGGTTGACACGTGTCAACGGATCCCAATTGCAATTAG
ATCGATGCTCATGCTGTTGATCCCTGGATTGCCTGGTCCGGACTCCACATACCGCCGCCATATTGC
TGGTATATTGTCGAATGCGACGCTAAACAGCAAGCCAACCAACAATACTGAGACCTGGTGGTACAT
GTCAGTTCTCTATTGCTGGGGTTCCAAACATAGCCATCAGTTGAAATATTTCATACATAGAAGAAT
ACCTCTGAATATGATGATGAAACATTTACTTAGACTTGCCTGTGAGCCCCAGGCAAAATGCACTGT
AAAAATACACTGACAGAGGATTAGGCATTCTTGGGAGTACTGTATAGTTAGTTGCATACATATTAG
CGTTCCCTCACTAAAACGAATCTCTGAATGCTATCAATTAAAGATCATGATGCTTTGATTGTGTCT
ACTGTATTTAAAATGGTGTTAAGATTTGCAATTACAATATACACAAACACGTTTCCTGCATCTCGG
AGAATGCAATCTTTCGTTGTACGCGTCTGTTTTCATATTTTTATGCATGTAGTTTGCACTACTTAG
CGTCCAATAAATCCATTCACAAAATCACACAAACAAACGATTTTAGGAATGTGACTGTAGCTGCAA
CGAATATACCTGATCCTTTCTTGTTCCAG

DOMAIN 2F-2 (2nd part of domain f)

ATCGCATGTTGCATTCACGGTATGCCGACCTTCCCCCAGTGGCACAGACTGTACACCCTGCAGTTG
GAGATGGCTCTGAGGAGACATGGATCATCTGTCGCCATCCCCTACTGGGACTGGACAAAGCCTATC
TCCGAACTCCCCTCGCTCTTCACCAGCCCTGAGTATTATGACCCATGGCATGATGCTGTGGTAAAC
AACCCATTCTCCAAAGGTTTTGTCAAATTTGCAAATACCTACACAGTAAGAGACCCACAGGAGATG
CTGTTCCAGCTTTGTGAACATGGAGAGTCAATCCTCTATGAGCAAACTCTTCTTGCTCTAGAGCAA
ACCGACTACTGTGATTTTGAGGTACAGTTTGAGGTCCTCCATAACGTGATCCACTACCTTGTTGGC
GGACGTCAGACCTACGCATTGTCTTCTCTGCATTATGCATCCTACGACCCATTCTTCTTTATACAC
CATTCCTTTGTGGATAAGATGTGGGTAGTATGGCAAGCTCTTCAAAAGAGGAGGAAACTTCCATAC
AAGCGAGCTGACTGTGCTGTCAACCTAATGACTAAACCAATGAGGCCATTTGACTCCGATATGAAT
CAGAACCCATTCACAAAGATGCACGCAGTTCCCAACACACTCTATGACTACGAGACACTGTACTAC
AGCTACGATAATCTCGAAATAGGTGGCAGGAATCTCGACCAGCTTCAGGCTGAAATTGACAGAAGC
AGAAGCCACGATCGCGTTTTTGCTGGATTCTTGCTTCGTGGAATCGGAACTTCTGCTGATGTCAGG
TTTTGGATTTGTAGAAATGAAAATGACTGCCACAGGGGTGGAATAATTTTCATCTTAGGTGGAGCC
AAGGAAATGCCATGGTCATTTGACAGAAACTTCAAGTTTGATATCACCCATGTACTCGAGAAAGCT
GGCATTAGCCCAGAGGACGTGTTTGATGCTGAGGAGCCATTTTATATCAAGGTTGAGATCCATGCT
GTTAACAAGACCATGATACCATCGTCTGTGATCCCAGCCCCAACTATCATCTATTCTCCTGGGGAA
G

INTRON 2F-2/2G-1 (SEQ ID NO:134)

GTGAGAGAACCAGTAATAGCTACTGTCTACAAAGAATGTGTTCATTTAAAGACCTGACTGTAGGCC
GATGGCTGCTGTCATCTCCTCCGCCTCCTCCTCCTGTTCCTCCTCCGAAGGGGTCAGCTTCAGGTT

CTCTTGCCAATATGCCAAGCAGACCTCCTGAGCAGGCAGTATATATACGTAAGGGAAGCAAGTATG
GACCATCGCGCGGCATGTAGAGATACAATGATCAGCTGTCTGCTGTTCCACTCCTGTCAGACAATG
AGATAAACATGAATACAGTATTACTCAGCAGCGTTCCAATTTTCAACCCTCGTATTTATTAAAAAA
AGGAATTTTTAATATATTTTTCTCCTTGTTGAAATATTTTAGTAACTGTTAATCGATATAGAGTGG
AGTAGTGACGCTTTATTTCGGTTCATTCTCGAAACAAAAATATAATAGTCCACTGAACTCTCTTAA
ATTGTTTTTACAACCTTCAACTGCCACAGACGTAATCCCTCACGTTATTTTGAGCTGACAACGTGT
TGAATTGAGTGTGTTCCGAATTCTAAATAAGCATGTATATATTTACGTCTCATGCAAGTAATATAT
GTTTAACTGATGACGTCACTTGGTGACCACTGATTTAGTTCCTTTGTCATAATTGCAGTTTCTGTT
GTCACGGGGACGGTGGGGAAGCCAGGTTCCTCCTGTCACGCTGAATATCCCGTTCGAATCCCCCAC
ATGGGTACAAAGTGTGATGCCTATTTCTGGTGTCCCCCACCGTGATATTGCTGGAATAAGTGGCTT
AATACCATATACACTCACTCTATTGTCACACTACTGCCACCGGCTCACACCTCTGATGCTTCTGTT
CTATCCAG

DOMAIN 2G-1 (1st part of domain g)

GTCGCGCTGCTGACAGTGCACACTCAGCCAACATTGCTGGCTCTGGGGTGAGGAAGGACGTCACGA
CCCTCACTGTGTCTGAGACCGAGAACCTAAGACAGGCTCTTCAAGGTGTCATCGATGATACTGGTC
CCAATGGTTACCAAGCAATAGCATCCTTCCACGGAAGTCCTCCAATGTGCGAGATGAACGGCCGCA
AGGTTGCCTGTTGTGCTCACG

INTRON 2G-1/2G-2 (SEQ ID NO:135)

GTAATTAATGGATGTGAAGTCAATGTCCGAGGGTATAATAAGGATTTAAATACTTCAGTCGTGTAA
TACTGTATGACATGTGTATTGGATGGTGTAGGTATTACAGGTTATAAGGCCAGTGTGTGTTGGGAC
GGTTACTTTCCTGCACTAGTAATAAGCATTGTATTTAGCTAGCTTTTATCATATAACTTTAGTTTC
ATGGTTTGTGGCAATTGAAATCGAAATTTTCTTTCATTTCAAGGTTATCGCACTCGTGTGTTAGAA
TAGTTACTATGCTGCATTGAGAATAACACTATAGTAATAAAGCATATCATACAGTAAGAATAACAC
TATAGTAATAAAGTATATCATACAGTAAGAATGTCATTGTATGATAAATAGGTTATCACACTCGTG
TGTTTTAGAATGGTTACTATCCCAGGAATAACCACTATGTATTACATGTATATTGGGCAGTGTAAG
TAGTAGCATTGTATATTAAATCAGTATATCGTGCTTCAAAACACCAGGATATATGGGGTATACAGT
GGGCAGTGTAAGTAGCAACATTGTATATTAAATCAGTATATCGTACTTCAAAACACCAGGATTATG
GGGTATACAGTGGGCAGTGTAAGTAGTAGCATTGTATATTAAATCAGTATATCGTACTTCAAAACA
CCAGGATATAATTCAGTATATCGTGCTTCAAAACACCAGGATATAATTCAGTATATCGTGCTTCAA
AACACCAGGATATATGGGATATACAGTGCGGGTTTGCATACAACCTCCACCCTTTACAG

DOMAIN 2G-2 (2nd part of domain g)

GTATGGCCTCCTTCCCACACTGGCACAGACTGTATGTGAAGCAGATGGAAGACGCCCTGGCTGACC
ACGGATCACATATCGGCATCCCTTACTGGGACTGGACAACTGCCTTCACAGAGTTACCCGCCCTTG
TCACAGACTCCGAGAACAATCCCTTCCATGAG

INTRON 2G-2/2G-3 (SEQ ID NO:136)

GTCAGTTTAGTCTCCTGTCTGAGCTAACGATACCAATTTCCTATTTTCGAGAACCACGATGACGAG
AAAACAAGCAATATAGATATAGATGCAGTATAGATCAAGTTAATGAATTCATTGCTATATGTTTGC
TTGTAATAAACTTTAAGAAAACGAGAGCATGCACACAAATGAAACAAACAATTATGTGTTTGATAG
GAATATGATATATGTATTTGGGGGCTGACGTGAGCAGGGTTGAAGGGACAGTTTACATTGTCAGTA
ACACTGGGAGTATTCTTTGATCCACAATATATAGTTTCATTGTGTTCAGCAGTTACAACTAACATT
ATATCATACATTACGTCGTAACATGCTTCTTTTGTCCTCTTCTGCCAG

DOMAIN G-3 (3rd part of domain g)

GGTCGCATTGATCATCTCGGTGTAACCACGTCACGTTCCCCCAGAGACATGCTGTTTAACGACCCA
GAGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTCCTGGCTTTGGAGCAGACTGACTACTGC

CAGTTCGAAGTCCAGTTTGAGCTGACCCACAACGCCATTCACTCCTGGACAGGTGGACGTAGCCCT
TACGGAATGTCGACCCTCGAGTTCACAGCCTACGATCCTCTCTTCTGGCTTCACCACTCCAACACC
GACAGAATCTGGGCTGTCTGGCAAGCACTGCAGAAATACCGAGGACTCCCATACAACGAAGCACAC
TGTGAAATCCAGGTTCTGAAACAGCCCTTGAGGCCATTCAACGATGACATCAACCACAATCCAATC
ACCAAGACTAATGCCAGGCCTATCGATTCATTTGATTATGAGAGGTTTAACTATCAGTATGACACC
CTTAGCTTCCATGGTAAGAGCATCCCTGAACTGAATGACCTGCTCGAGGAAAGAAAAAGAGAAGAG
AGAACATTTGCTGCCTTCCTTCTTCGTGGAATCGGTTGCAGTGCTGATGTCGTCTTTGACATCTGC
CGCCCCAATGGTGACTGTGTCTTTGCAGGAACCTTTGCTGTGCTGGGAGGGGAGCTAGAAATGCCT
TGGTCCTTCGACAGACTGTTCCGCTATGACATCACCAGAGTCATGAATCAGCTCCATCTCCAGTAT
GATTCAGATTTCAGTTTCAGGGTGAAGCTTGTTGCAACCAATGGCACTGAGCTTTCATCAGACCTC
CTCAAGTCACCAACAATTGAACATGAACTTGGAG

INTRON 2G-3/2H (SEQ ID NO:137)

GTATGTTATCTTATTATCAAATGTGTAATCAGATACTGGAGACGTTTTCATATTAACTTGGTCAGC
ATTAGTTGATGATTTTGGTGCGATATTGACGACAAGGAGTTAAGCATTAACACGTTCAACACATCT
TTAATCTGATATGAGAAGGGAATAAATTGATCCAGTATTGATGATTGAAGTTAGATTAACAGTGAA
AGATATACCAGTTTTGATAATCGTATAAAACAGTAGCAGAATTGTATCGTGAAAACTAAATGTGGG
AAGGCGAACGCCAAGCAGATTTTAGATTACGATCGTGTGCTAGAATAATTCACAATAACCCAGACG
TCGGAAATGTGGTTGTCTATGGCAATAGTTACGATTAATTGCTAACATGCACGATTTACCTATTTC
AG

DOMAIN 2H

CCCACAGAGGACCAGTTGAAGAAACAGAAGTCACTCACCAAAATACTGACGGCAATGCACACTTCC
ATCGTAAGGAAGTTGATTCGCTGTCCCTGGATGAAGCAAACAACTTGAAGAATGCCCTTTACAAGC
TACAGAACGACCACAGTCTAACAGGATACGAAGCAATCTCTGGTTACCATGGATACCCGAATCTGT
GTCCGGAAGAAGGCGATGACAAATACCCCTGCTGCGTCCACGGAATGGCCATCTTCCCCCACTGGC
ACAGACTCTTGACCATCCAACTGGAAAGAGCTCTCGAGCACAATGGTGCACTGCTTGGTGTTCCTT
ACTGGGACTGGACCAAGGACCTGTCGTCACTGCCGGCGTTCTTCTCCCGACTCCAGCAACAACAATC
CCTACTTCAAGTACCACATCGCAGGTGTTGGTCACGACACCGTCAGAGAGCCAACTAGTCTTATAT
ATAACCAGCCCCAAATCCATGGTTATGATTATCTCTATTACCTAGCATTGACCACGCTTGAAGAAA
ACAATTACTGTGACTTTGAGGTTCAGTATGAGATCCTCCACAACGCCGTCCACTCCTGGCTTGGAG
GATCCCAGAAGTATTCCATGTCTACCCTGGAGTATTCGGCCTTTGACCCTGTCTTTATGATCCTTC
ACTCGGGTCTAGACAGACTTTGGATCATCTGGCAAGAACTTCAGAAGATCAGGAGAAAGCCCTACA
ACTTCGCTAAATGTGCTTATCATATGATGGAAGAGCCACTGGCGCCCTTCAGCTATCCATCTATCA
ACCAGGACGAGTTCACCCGTGCCAACTCCAAGCCTTCTACAGTTTTTGACAGCCATAAGTTCGGCT
ACCATTACGATAACCTGAATGTTAGAGGTCACAGCATCCAAGAACTCAACACAATCATCAATGACT
TGAGAAACACAGACAGAATCTACGCAGGATTTGTTTTTGTCAGGCATCGGTACGTCTGCTAGTGTCA
AGATCTATCTCCGAACAGATGACAATGACGAAGAAGTTGGAACTTTCACTGTCCTGGGAGGAGAGA
GGGAAATGCCATGGGCCTACGAGCGAGTTTTCAAGTATGACATCACAGAGGTTGCAGATAGACTTA
AACTAAGTTATGGGGACACCTTTAACTTCCGACTAGAGATCACATCCTACGATGGATCGGTGGTAA
ACAAGAGCCTACCCAATCCTTTCATCATCTACAGACCTGCCAATCATGACTACGATGTTCTTGTTA
TCCCAGTAGGAAGAAACCTTCACATCCCTCCCAAAGTTGTCGTCAAGAGAGGCACCCGCATCGAGT
TCCACCCAGTCGATGATTCAGTTACGAGACCAGTTGTTGATCTTGGAAGCTACACTGCACTCTTCA
ACTGTGTGGTACCACCGTTCACATACCGCGGATTCGAACTGAACCACGTCTATTCTGTCAAGCCTG
GTGACTACTATGTTACCGGACCAACGAGAGACCTTTGCCAGAATGCAGATGTCAGGATTCATATCC
ATGTTGAGGATGAGTAA

3′UTR

CGCAACAG

INTRON 3´UTR (SEQ ID NO:138)

GTGAGATAAGAAACCCTTCTAACAGTAATACGACACCACATTACAGCTTAAACATGATTGCCATCG
ATGTTTTCATGTGTAGTATACGCTTTTCAGTTCTACATAATTTTGTTTTTCAAATCAAGTTTAGCA
AATGAATCTATCACTGGAAAATAGGGTAGGGTAGCCAAGTGGTTAAAGCGGTCACTGATCACGCCA
AAGACGAGTGTCCTAACCTGCATGGGTACAAAAGTGAAGACCATTGCTGGTGTCTACCGCCGTAAT
ATTGTTTTTAGTATTGCTAAAACTTATACTCACCCATGCGCTGTAAAAGTGGAATAATAATCATAT
TTCAACAAAAGCACAAAACCATTTCATTTTCATGAAAGCCTCTTGTTCACCTGAAAGACGCAAGAG
AACAATAGTTCCTAACATTATTTTCAGACATTGGAAATGTCCTGCACGTGTAAACCATATATCCTT
TGAAATTTTTACGACTGCATCGTATACAATTTATGATATAAATTTAAAACTTTATTTCAG

3´UTR

GTTTCTTGGTCTCCACATATTCACACATCAGCACCAAACGGTTTCGAAGGACATTGGCGTTCTTCT
CTGGCAATGCATTTCAATACAACATTGAAAATGACTTCAGCATATCAGTGTGCTTCGAACGTGTTC
CGGAAGTACTCAAATGTGCTATGACTGAATTATTGTACATACATAACTTATTGATGTTCAATAAAT
AAATGTTGAAACG

Figure 7

Primary structure of the HtH2 protein

DOMAIN A (SEQ ID NO:156)

GLPYWDWTQHLTQLPDLVSDPLFVDPEGGKAHDNAWYRGNIKFENKKTARAVDDRLFEKVGPGENT
RLFEGILDALEQDEFCNFEIQFELAHNAIHYLVGGRHTYSMSHLEYTSYDPLFFLHHSNTDRIFAI
WQRLQVLRGKDPNTADCAHNLIHEPMEPFRRDSNPLDLTRENSKPIDSFDYAHLGYQYDDLTLNGM
TPEELNSYLHERSGKEGVFASFRLSGFGGSANVVVYACRPAHDEMAVDQCDKAGDFFVLGGPTEMP
WRFYRAFHFDVTDSIDNIDKDRHGHYYVKAELFSVNGSALPNDLLPQPTISHRPARGHVDEAPAPS
SDAHLAVRKDINHLTREEVYELRRAMERFQADTSVDGYQATVEYHGLPARCPFPEATNRFACCIHG
MATFPHW

DOMAIN B

HRLFVTQVEDALIRRGSPIGVPYWDWTQPMAHLPGLADNATYRDPISGDSRHNPFHDVEVAFENGR
TERHPDSRLFEQPLFGKHTRLFDSIVYAFEQEDFCDFEVQFEMTHNNIHAWIGGGGKYSMSSLHYT
AFDPISYLHHSNTDRLWAIWQALQIRRNKPYKAHCAWSEERQPLKPFAFSSPLNNNEKTYENSVPT
NVYDYEGVLGYTYDDLNFGGMDLGQLEEYIQRQRQRDRTFAGFFLSHIGTSANVEIIIDHGTLHTS
VGTFAVLGGEKEMKWGFDRLYKYEITDELRQLNLRADDGFSISVKVTDVDGSELSSELIPSAAIIF
ERSH

DOMAIN C

IDHQDPHQDTIIRKNVDNLTPEEINSLRRAMADLQSDKTAGGFQQIAAFHGEPKWCPSPDAEKKFS
CCVHGMAVFPHWHRLLTVQGENALRKHGCLGALPYWDWTRPLSHLPDLVSQQNYTDAISTVEARNP
WYSGHIDTVGVDTTRSVRQELYEAPGFGHYTGVAKQVLLALEQDDFCDFEVQFEIAHNFIHALVGG
SEPYGMASLRYTTYDPIFYLHHSNTDRLWAIWQALQKYRGKPYNSANCAIASMRKPLQPFGLTDEI
NPDDETRQHAVPFSVFDYKNNFNYEYDTLDFNGLSISQLDRELSRRKSHDRVFAGFLLHGIQQSAL
VKFFVCKSDDDCDHYAGEFYILGDEAEMPWGYDRLYKYEITEQLNALDLHIGDRFFIRYEAFDLHG
TSLGSNIFPKPSVIHDEGA

DOMAIN D

GHHQADEYDEVVTAASHIRKNLKDLSKGEVESLRSAFLQLQNDGVYENIAKFHGKPGLCDDNGRKV
ACCVHGMPTFPQWHRLYVLQVENALLERGSAVSVPYWDWTETFTELPSLIAEATYFNSRQQTFDPN
PFFRGKISFENAVTTRDPQPELYVNRYYYQNVMLAFEQDNYCDFEIQFEMVHNVLHAWLGGRATYS
ISSLDYSAFDPVFFLHHANTDRLWAIWQELQRYRKKPYNEADCAINLMRKPLHPFDNSDLNHDPVT
FKYSKPTDGFDYQNNFGYKYDNLEFNHFSIPRLEEIIRIRQRQDRVFAGFLLHNIGTSATVEIFVC
VPTTSGEQNCENKAGTFAVLGGETEMAFHFDRLYRFDISETLRDLGIQLDSHDFDLSIKIQGVNGS
YLDPHILPEPSLIFVPGSS

DOMAIN E

SFLRPDGHSDDILVRKEVNSLTTRETASLIHALKSMQEDHSPDGFQAIASFHALPPLCPSPSATHR
YACCVHGMATFPQWHRLYTVQFQDALRRHGAAVGVPYWDWLRPQSHLPELVTMETYHDIWSNRDFP
NPFYQANIEFEGENITTEREVIADKLFVKGGHVFDNWFFKQAILALEQENYCDFEIQFEILHNGVH
TWVGGSRTHSIGHLHYASYDPLFYLHHSQTDRIWAIWQELQEQRGLSGDEAHCALEQMREPLKPFS
FGAPYNLNQLTQDFSRPEDTFDYRKFGYEYDNLEFLGMSVAELDQYIIEHQENDRVFAGFLLSGFG
GSASVNFQVCRADSTCQDAGYFTVLGGSAEMAWAFDRLYKYDITETLEKMHLRYDDDFTISVSLTA
NNGTVLSSSLIPTPSVIFQRGH

DOMAIN F

RDINTKSMSANRVRRELSDLSARDPSSLKSALRDLQEDDGPNGYQALAAFHGLPAGCHDSQGNEIA
CCIHGMPTFPQWHRLYTLQLEMALRRHGSSVAIPYWDWTKPISELPSLFTSPEYYDPWHDAVVNNP
FSKGFVKFANTYTVRDPQEMLFQLCEHGESILYEQTLLALEQTDYCDFEVQFEVLHNVIHYLVGGR
QTYALSSLHYASYDPFFFIHHSFVDKMWVVWQALQKRRKLPYKRADCAVNLMTKPMRPFDSDMNQN
PFTKMHAVPNTLYDYETLYYSYDNLEIGGRNLDQLQAEIDRSRSHDRVFAGFLLRGIGTSADVRFW
ICRNENDCHRGGIIFILGGAKEMPWSFDRNFKFDITHVLEKAGISPEDVFDAEEPFYIKVEIHAVN
KTMIPSSVIPAPTIIYSPGE

DOMAIN G

GRAADSAHSANIAGSGVRKDVTTLTVSETENLRQALQGVIDDTGPNGYQAIASFHGSPPMCEMNGR
KVACCAHGMASFPHWHRLYVKQMEDALADHGSHIGIPYWDWTTAFTELPALVTDSENNPFHEGRID
HLGVTTSRSPRDMLFNDPEQGSESFFYRQVLLALEQTDYCQFEVQFELTHNAIHSWTGGRSPYGMS
TLEFTAYDPLFWLHHSNTDRIWAVWQALQKYRGLPYNEAHCEIQVLKQPLRPFNDDINHNPITKTN
ARPIDSFDYERFNYQYDTLSFHGKSIPELNDLLEERKREERTFAAFLLRGIGCSADVVFDICRPNG
DCVFAGTFAVLGGELEMPWSFDRLFRYDITRVMNQLHLQYDSDFSFRVKLVATNGTELSSDLLKSP
TIEHEL

DOMAIN H

GAHRGPVEETEVTHQNTDGNAHFHRKEVDSLSLDEANNLKNALYKLQNDHSLTGYEAISGYHGYPN
LCPEEGDDKYPCCVHGMAIFPHWHRLLTIQLERALEHNGALLGVPYWDWTKDLSSLPAFFSDSSNN
NPYFKYHIAGVGHDTVREPTSLIYNQPQIHGYDYLYYLALTTLEENNYCDFEVQYEILHNAVHSWL
GGSQKYSMSTLEYSAFDPVFMILHSGLDRLWIIWQELQKIRRKPYNFAKCAYHMMEEPLAPFSYPS
INQDEFTRANSKPSTVFDSHKFGYHYDNLNVRGHSIQELNTIINDLRNTDRIYAGFVLSGIGTSAS
VKIYLRTDDNDEEVGTFTVLGGEREMPWAYERVFKYDITEVADRLKLSYGDTFNFRLEITSYDGSV
VNKSLPNPFIIYRPANHDYDVLVIPVGRNLHIPPKVVVKRGTRIEFHPVDDSVTRPVVDLGSYTAL
FNCVVPPFTYRGFELNHVYSVKPGDYYVTGPTRDLCQNADVRIHIHVEDE

Figure 8

Genomic sequence of the KLH1 gene

DOMAIN 1B

GGCCTACCGTACTGGGACTGGACTGAACCCATGACACACATTCCGGGTCTGGCAGGAAACAAAACT
TATGTGGATTCTCATGGTGCATCCCACACAAATCCTTTTCATAGTTCAGTGATTGCATTTGAAGAA
AATGCTCCCCACACCAAAAGACAAATAGATCAAAGACTCTTTAAACCCGCTACCTTTGGACACCAC
ACAGACCTGTTCAACCAGATTTTGTATGCCTTTGAACAAGAAGATTACTGTGACTTTGAAGTCCAA
TTTGAGATTACCCATAACACGATTCACGCTTGGACAGGAGGAAGCGAACATTTCTCAATGTCGTCC
CTACATTACACAGCTTTCGATCCTTTGTTTTACTTTCACCATTCTAACGTTGATCGTCTTTGGGCC
GTTTGGCAAGCCTTACAGATGAGACGGCATAAACCCTACAGGGCCCACTGCGCCATATCTCTGGAA
CATATGCATCTGAAACCATTCGCCTTTTCATCTCCCCTTAACAATAACGAAAAGACTCATGCCAAT
GCCATGCCAAACAAGATCTACGACTATGAAAATGTCCTCCATTACACATACGAAGATTTAACATTT
GGAGGCATCTCTCTGGAAAACATAGAAAAGATGATCCACGAAAACCAGCAAGAAGACAGAATATAT
GCCGGTTTTCTCCTGGCTGGCATACGTACTTCAGCAAATGTTGATATCTTCATTAAAACTACCGAT
TCCGTGCAACATAAGGCTGGAACATTTGCAGTGCTCGGTGGAAGCAAGGAAATGAAGTGGGGATTT
GATCGCGTTTTCAAGTTTGACATCACGCACGTTTTGAAAGATCTCGATCTCACTGCTGATGGCGAT
TTCGAAGTTACTGTTGACATCACTGAAGTCGATGGAACTAAACTTGCATCCAGTCTTATTCCACAT
GCTTCTGTCATTCGTGAGCATGCACGTGGTAAGCTGAATAGAG

INTRON 1B/1C(SEQ ID NO:139)

GTTTTGTAATAATTATGTAGAATTCTTTACCTCAGAATAAGATGAGGTCACATGGGTTTTGCAAAA
CTATTACGTTCGAATTAATATTAATAATACCGGACCCTCCACTGGTACATATTTATCTTTATAACG
ATAATAGCGATGATGATGATGATGATGATGATGATGATGATGATGATAATGATGATGCCGGTATTG
CACGTAATCCAGCCGACTTAGATGACACCCTAAGGGTGCAGAAAGTATAACAATTAGATTGCGTTT
GCATCTGTGTATGCGTGTGCTTTAACCAAAAGTCAAAATAAAAGTGCAAACCCTTAGTTTATTCAT
TTGATAGAGCCTTTTACGATAAGAACAATGTAATAAATTAGAACATAACTGAAACCTCCGAAAGAA
GGCCTGTTTGTCAAGAGAGGTATCGACATGATTGACTTATAAACCTGTGCTTCTATATTTTGGAAC
TGTCCACTTTCTTGTTGTGTGTACTGTAATCACATCGCACTATGGCTGCAAGACGTGTACGAGTAC
ACTATATACTTACCTAATGACCAACCACAAGGCTGGCTTTGTTAATATTGTTATTTCACAGAAATA
AACACAGAATTCCAGCATTTGGCTGGTGTATTTAGCAAAACACCGATATGACACTCATGTTTTATT
ACATTTTTTTCAG

DOMAIN 1C

TTAAATTTGACAAAGTGCCAAGGAGTCGTCTTATTCGAAAAAATGTAGACCGTTTGAGCCCCGAGG
AGATGAATGAACTTCGTAAAGCCCTAGCCTTACTGAAAGAGGACAAAAGTGCCGGTGGATTTCAGC
AGCTTGGTGCATTCCATGGGGAGCCAAAATGGTGTCCTAGTCCCGAAGCATCTAAAAAATTTGCCT
GCTGTGTTCACGGCATGTCTGTGTTCCCTCACTGGCATCGACTGTTGACGGTTCAGAGTGAAAATG
CTTTGAGACGACATGGCTACGATGGAGCTTTGCCGTACTGGGATTGGACCTCTCCTCTTAATCACC
TTCCCGAACTGGCAGATCATGAGAAGTACGTCGACCCTGAAGATGGGGTAGAGAAGCATAACCCTT
GGTTCGATGGTCATATAGATACAGTCGACAAAACAACAAGAAGTGTTCAGAATAAACTCTTCG
AACAGCCTGAGTTTGGTCATTATACAAGCATTGCCAAACAAGTACTGCTAGCGTTGGAACAGGACA
ATTTCTGTGACTTTGAAATCCAATATGAGATTGCCCATAACTACATCCATGCACTTGTAGGAGGCG
CTCAGCCTTATGGTATGGCATCGCTTCGCTACACTGCTTTTGATCCACTATTCTACTTGCATCACT
CTAATACAGATCGTATATGGGCAATATGGCAGGCTTTACAGAAGTACAGAGGAAAACCGTACAACG
TTGCTAACTGTGCTGTTACATCGATGAGAGAACCTTTGCAACCATTTGGCCTCTCTGCCAATATCA
ACACAGACCATGTAACCAAGGAGCATTCAGTGCCATTCAACGTTTTTGATTACAAGACCAATTTCA
ATTATGAATATGACACTTTGGAATTTAACGGTCTCTCAATCTCTCAGTTGAATAAAAAGCTCGAAG

CGATAAAGAGCCAAGACAGGTTCTTTGCAGGCTTCCTGTTATCTGGTTTCAAGAAATCATCTCTTG
TTAAATTCAATATTTGCACCGATAGCAGCAACTGTCACCCCGCTGGAGAGTTTTACCTTCTGGGTG
ATGAAAACGAGATGCCATGGGCATACGATAGAGTCTTCAAATATGACATAACCGAAAAACTCCACG
ATCTAAAGCTGCATGCAGAAGACCACTTCTACATTGACTATGAAGTATTTGACCTTAAACCAGCAA
GCCTGGGAAAAGATTTGTTCAAGCAGCCTTCAGTCATTCATGAACCAAGAATAG

INTRON 1C/1D (SEQ ID NO:140)

GTACTTGTTATATGTTTCGAATATTGCCGATACCTTCAATATATATACTTTATCAAAGTAATTGAT
TAATCTGAAGTAATTTTCCTTTCCAGTAGAGATTCAGTTGATACAACAAGAATTCGCCCTGTTGTA
TGTCACTTTATTTTCATCAAACGATTCGAAGTGAGCTGTCCATGCCACAATGGGGTCTCTGTAACT
TTCTCGTATGGGGTATAGATTATATAGACGTGGCAGACCTTACGTATAACTAATATTTGTGTAATG
TCGTTTCAG

DOMAIN 1D

GTCACCATGAAGGCGAAGTATATCAAGCTGAAGTAACTTCTGCCAACCGTATTCGAAAAAACATTG
AAAATCTGAGCCTTGGTGAACTCGAAAGTCTGAGAGCTGCCTTCCTGGAAATTGAAAACGATGGAA
CTTACGAATCAATAGCTAAATTCCATGGTAGCCCTGGTTTGTGCCAGTTAAATGGTAACCCCATCT
CTTGTTGTGTCCATGGCATGCCAACTTTCCCTCACTGGCACAGACTGTACGTGGTTGTCGTTGAGA
ATGCCCTCCTGAAAAAAGGATCATCTGTAGCTGTTCCCTATTGGGACTGGACAAAACGAATCGAAC
ATTTACCTCACCTGATTTCAGACGCCACTTACTACAATTCCAGGCAACATCACTATGAGACAAACC
CATTCCATCATGGCAAAATCACACACGAGAATGAAATCACTACTAGGGATCCCAAGGACAGCCTCT
TCCATTCAGACTACTTTTACGAGCAGGTCCTTTACGCCTTGGAGCAGGATAACTTCTGTGATTTCG
AGATTCAGTTGGAGATATTACACAATGCATTGCATTCTTTACTTGGTGGCAAAGGTAAATATTCCA
TGTCAAACCTTGATTACGCTGCTTTTGATCCTGTGTTCTTCCTTCATCACGCAACGACTGACAGAA
TCTGGGCAATCTGGCAAGACCTTCAGAGGTTCCGAAAACGGCCATACCGAGAAGCGAATTGCGCTA
TCCAATTGATGCACACGCCACTCCAGCCGTTTGATAAGAGCGACAACAATGACGAGGCAACGAAAA
CGCATGCCACTCCACATGATGGTTTTGAATATCAAAACAGCTTTGGTTATGCTTACGATAATCTGG
AACTGAATCACTACTCGATTCCTCAGCTTGATCACATGCTGCAAGAAAGAAAAAGGCATGACAGAG
TATTCGCTGGCTTCCTCCTTCACAATATTGGAACATCTGCCGATGGCCATGTATTTGTATGTCTCC
CAACTGGGGAACACACGAAGGACTGCAGTCATGAGGCTGGTATGTTCTCCATCTTAGGCGGTCAAA
CGGAGATGTCCTTTGTATTTGACAGACTTTACAAACTTGACATAACTAAAGCCTTGAAAAAGAACG
GTGTGCACCTGCAAGGGGATTTCGATCTGGAAATTGAGATTACGGCTGTGAATGGATCTCATCTAG
ACAGTCATGTCATCCACTCTCCCACTATACTGTTTGAGGCCGGAACAG

INTRON 1D/1E (SEQ ID NO:141)

GTAACTATTTTGTCACTGTAACCAACAACTGCAGTCTATTTTGCAATTACGATAATAACAATTTTT
GAAATATATCTTTATTAAAGCAAAGGTTTCTAGAGACAAACAGCCGGCTCTAATTATTTTTTCGAA
CTTACGCTTGAGTAAAGATCTGCAAATGGCAACCCTACCTATACTATTAAAAATATAATGTTACAT
TCGTATCTGAATGTTTAATAAATCACTTCATATTCTGTTGCAG

DOMAIN 1E

ATTCTGCCCACACAGATGATGGACACACTGAACCAGTGATGATTCGCAAAGATATCACACAATTGG
ACAAGCGTCAACAACTGTCACTGGTGAAAGCCCTCGAGTCCATGAAAGCCGACCATTCATCTGATG
GGTTCCAGGCAATCGCTTCCTTCCATGCTCTTCCTCCTCTTTGTCCATCACCAGCTGCTTCAAAGA
GGTTTGCGTGCTGCGTCCATGGCATGGCAACGTTCCCACAATGGCACCGTCTGTACACAGTCCAAT
TCCAAGATTCTCTCAGAAAACATGGTGCAGTCGTTGGACTTCCGTACTGGGACTGGACCCTACCTC
GTTCTGAATTACCAGAGCTCCTGACCGTCTCAACTATTCATGACCCGGAGACAGGCAGAGATATAC
CAAATCCATTTATTGGTTCTAAAATAGAGTTTGAAGGAGAAAACGTACATACTAAAAGAGATATCA
ATAGGGATCGTCTCTTCCAGGGATCAACAAAAACACATCATAACTGGTTTATTGAGCAAGCACTGC
TTGCTCTTGAACAAACCAACTACTGCGACTTCGAGGTTCAGTTTGAAATTATGCATAATGGTGTTC

ATACCTGGGTTGGAGGCAAGGAGCCCTATGGAATTGGCCATCTGCATTATGCTTCCTATGATCCAC
TTTTCTACATCCATCACTCCCAAACTGATCGTATTTGGGCTATATGGCAATCGTTGCAGCGTTTCA
GAGGACTTTCTGGATCTGAGGCTAACTGTGCTGTAAATCTCATGAAAACTCCTCTGAAGCCTTTCA
GCTTTGGAGCACCATATAATCTTAATGATCACACGCATGATTTCTCAAAGCCTGAAGATACATTCG
ACTACCAAAAGTTTGGATACATATATGACACTCTGGAATTTGCAGGGTGGTCAATTCGTGGCATTG
ACCATATTGTCCGTAACAGGCAGGAACATTCAAGGGTCTTTGCCGGATTCTTGCTTGAAGGATTTG
GCACCTCTGCCACTGTCGATTTCCAGGTCTGTCGCACAGCGGGAGACTGTGAAGATGCAGGGTACT
TCACCGTGTTGGGAGGTGAAAAAGAAATGCCTTGGGCCTTTGATCGGCTTTACAAGTACGACATAA
CAGAAACCTTAGACAAGATGAACCTTCGACATGACGAAATCTTCCAGATTGAAGTAACCATTACAT
CCTACGATGGAACTGTACTCGATAGTGGCCTTATTCCCACACCGTCAATCATCTATGATCCTGCTC
ATC

INTRON 1E/1F (SEQ ID NO:142)

GTAAGTATACACACATTATTTCTCTTCTGCTATATCAGATGAAGAGAACGTTGTATCACTAACCTA
GTCTTGTTTGATTTGTGGTTTCGTTTGCTTCCTGAACAGTAGGGTTGATTTAACTTCTCTGTTTCG
TCTGTACCAATGAAAGACTATGATGCTTGTGTGAAGATGCTTTGTTCATGAGTCAGTCTGTTCTTG
TAATGCTTTGATCTTTGCCATCAACATTCTTGAAATTAATTATGGTTTCCCTTAAATACTTACATA
TTACATTTAAACGTCGCTGCTTGTCTGATTGCATATTCTTTCAAAAATAACTATATATTCCAG

DOMAIN 1F-1 (1st part of domain f)

ATGATATTAGTTCGCACCACCTGTCGCTCAACAAGGTTCGTCATGATCTGAGTACACTGAGTGAGC
GAGATATTGGAAGCCTTAAATATGCTTTGAGCAGCTTGCAGGCAGATACCTCAGCAGATGGTTTTG
CTGCCATTGCATCCTTCCATGGTCTGCCTGCCAAATGTAATGACAGCCACAATAACGAG

INTRON 1F-1/1F-2 (SEQ ID NO:143)

GTAAATATACAGTGAAATCCGGATAAGTAAAATCCAGATAAGAAAAAAAACATTTTCTGTGGTCCC
GGCATGTTTCTTCTTCATCTATCATTATTTTGATACGGATAAGTAAAAATCGGCTGAGTAAAACAT
CCGGGTAAGTAAAATGATTTTCGAGGTCTCTTCATCGGATAAGTAAGATACACAAGTGATCATTCC
AATAAACACTAACTGATGCAACACAATACCAGCGCACAGTGTTTTCACTACGTTTGTTTGTATTGT
AATTAACAATTAACACTTAAGTGTTTCCCAATGTGTCCGTGTGCAAACTGATTGGGACAAAGCTTG
CAACAAGCCCGGCAATTCCATGTCGTTTATGTCTACGTTTGTTATTCTGACTGCTTGGAGGGGTTC
GGAAAAAAATAAAAAACGGGTAAATATTATAAAAAATTCACGGTGCCTTGAAATTTTAGGTGTCCG
GATTTCACTGTAGATGATTAATTTCTCACTTGTAAACAAAAGGACCCCAGTACCCTCATTCGTGAC
GTACGTTATAAAATGTAATTATAAAAAGCCCATTATCATGTTATACGTGATCTTGNCTTGCAATTA
TNCTACCGCTTTCTTGATTTTTTAAAGCAATTTCTCCCTCTATGAACTTATTAACATAGCACTCCT
GCAAAAGAAAACAGTCACTGCATGGATCCATATTGAATGTTGCTGCTTATTTCTCATTTTATTACT
CACAGATATTTCAAGAACATCGTACTCTCTAACCAGGCTAAAGCAAAGAGGGTTACATTTTAGCCG
ACAAGTTCACTAGCTGAGTGGAACACGTATATATTAATGGAGATGACTCTGGTCATGATGATTAGG
ACAATTATCATGACGTTATCATTGATCATGACCATGTCAGTATAATAGATAGCTAACAAATAATGT
AATTACTAATTATGAAGCAATGGTGCATTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

GTGGCATGCTGTATCCATGGAATGCCTACATTCCCCCACTGGCACAGACTCTACACCCTCCAATTT
GAGCAAGCTCTAAGAAGACATGGCTCTAGTGTAGCAGTACCCTACTGGGACTGGACAAAGCCAATA
CATAATATTCCACATCTGTTCACAGACAAAGAATACTACGATGTCTGGAGAAATAAAGTAATGCCA
AATCCATTTGCCCGAGGGTATGTCCCCTCACACGATACATACACGGTAAGAGACGTCCAAGAAGGC
CTGTTCCACCTGACATCAACGGGTGAACACTCAGCGCTTCTGAATCAAGCTCTTTTGGCGCTGGAA
CAGCACGACTACTGCGATTTTGCAGTCCAGTTTGAAGTCATGCACAACACAATCCATTACCTAGTG
GGAGGACCTCAAGTCTATTCTTTGTCATCCCTTCATTATGCTTCATATGATCCGATCTTCTTCATA
CACCACTCCTTTGTAGACAAGGTTTGGGCTGTCTGGCAGGCTCTTCAAGAAAAGAGAGGCCTTCCA

TCAGACCGTGCTGACTGCGCTGTTAGTCTGATGACTCAGAACATGAGGCCTTTCCATTACGAAATT
AACCATAACCAGTTCACCAAGAAACATGCAGTTCCAAATGATGTTTTCAAGTACGAACTCCTGGGT
TACAGATACGACAATCTGGAAATCGGTGGCATGAATTTGCATGAAATTGAAAAGGAAATCAAAGAC
AAACAGCACCATGTGAGAGTGTTTGCAGGGTTCCTCCTTCACGGAATTAGAACCTCAGCTGATGTC
CAATTCCAGATTTGTAAAACATCAGAAGATTGTCACCATGGAGGCCAAATCTTCGTTCTTGGGGGG
ACTAAAGAGATGGCCTGGGCTTATAACCGTTTATTCAAGTACGATATTACCCATGCTCTTCATGAC
GCACACATCACTCCAGAAGACGTATTCCATCCCTCTGAACCATTCTTCATCAAGGTGTCAGTGACA
GCCGTCAACGGAACAGTTCTTCCGGCTTCAATCCTGCATGCACCAACCATTATCTATGAACCTGGT
CTCGGTG

INTRON 1F-2/1G-1 (SEQ ID NO:144)

GTCTCGGTGAGTTATTAAAAGAAACAAAATATTTACCATTACCATTGTTAACTACAAAAATGAGTG
AGATATCTTATATCACTGGTACACTACTGATATTTTATGCAATGAAATTACTATTTTTCCAGGTAC
GCTTCAACCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATCATGCTTTTCTGT
AAAACATAAAACACCAATTAACAATGTTCTTAGTGTGTTTGTTGACTCCCTTCCACTGCAACGCCT
ACATAATCAAAGTGTTCGTTTTTTTCCAAACTTTCCAGTTAGTGTTGAAGACTAAAAAGTTAAATA
AGCATTCACATAACTTCTAAGAGCAACTGGGACCATGCAGTTACGTATTGATATTTCTGTGAGAGT
GAAGCAAAACACTGTTTTTCAAGCTTAGGTTTATCAATCAAAATGTCCAATAGTTCATGTTATCGA
AAAGGCAGCGAAGGATAAGAGGCTCCGAGACATCTTGTCTATTCTCGTGTTCATATGATATCAACT
GAGGAGCTTCCATTACATTTTTGACCTTATCATTTAAAGACATACATGGAACATTTTCATTTTACA
GTTAAAGTGAACCACTTCAGGTTCAACTTCAACTTCGAATTCAACTTCTGTTGTGTGTTTTATGAG
CCGACTGAAATAGAGTGCCTTACTTTCACTTCTAGTTTCGTTCTGTCTCGTCATCGTTGTTTCTTT
CAGTGTGCATAGTACACGCCTAGTATAGAACACACGAACTTGTCCTTACTTAATAGATTCTGAAAC
TATTATGTGGAAAGTTGGCAGGCTATAGTAACATCCTGGCAAAATTATCATGTATCCTCTTGTTTG
TCATAATTAG

DOMAIN 1G-1 (1st part of domain g)

ACCATCACGAAGATCATCATTCTTCTTCTATGGCTGGACATGGTGTCAGAAAGGAAATCAACACAC
TTACCACTGCAGAGGTGGACAATCTCAAAGATGCCATGAGAGCCGTCATGGCAGACCACGGTCCAA
ATGGATACCAGGCTATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGA
ATTACTCGTGTTGTACACATG

INTRON G1-1/1G-2 (SEQ ID NO:145)

GTATGTATTTCCCACTGGTGGTCGCTGACTGCCAACACATACTTGTAATTTATTCATGAAAGTATA
ATAGTTTGTTTGAAAGTATATTTATAACCATCTTGCACAAGCGTCACGAATTTTCACCACAAAGCT
TCAAAACGCCCAAAACATTCTAATAGCGATATATTTGTTAAAAGACCAAAATATAGCCTTACAACA
ATAGATTATTTTAATAAGACCAGTCAGTGCATGCAAATCGATTGGAAACTTTGAAATAAAATATTC
TATGTACTAACTGCCAATCTCATAATACTTGCCTTGGATGTGCTTCTTTTTCACATTCGCGTCGAG
CTTCAACTCCAATGCATAAGCTTAAAAATAATCATAAACACAAACAAATAGCCACAGAGGCGACGA
TCCCTCCAGGCCAGGCTTTATTTGTCTCTTATAGAATATATCGCTATTAGAATGTTTTTGACGTTT
TGAAGCTTTGTGGGTGAAAATTCGTGATGTTTATGCGTGGTATTTATGTAAGATGAAAATAAATAT
ATCTTTTCAAACAAGATTTTAGTATTTTGAAGACTTCTATGAATAAATTACACTTATGTGTTAGGT
TATTGGTCACTGAGCGCTTGTGGTATTTTCCCTTCTTCAATTTGTTTGTTCTTTGTTCAATTTCGA
ATAGTTATCCTACTGTGGATAGTCTATATGAGAATCGTTGAAAGAATAATACAATTCTAATGGATT
GCAACTTCTTTAACTTTATTTGCAACTGCCACGTTTCGGTATACGTTCTTATGCCGTCATCAAGC
ATACGAGTGTACATGTATGCCAAAACGCTGCAAATAAAAATTAAAGAAGTTGCAATCCATAAGAAT
TTCAATGTTCTTTCATCATCACATCAACTTCTAAAAATGCCTATAAAACAATCAACAAACGTACAA
TAGTACATTACCGGATCTCGCAGCATGACCACGTCGATATCTAAACAATATCACTATCCATTAATA
GGATCAAGAGTAGGTACAGACATGTTCAGTTATAAATACTCTTCAAAAAAGTAGGGGAACTTGGAA
TTTCAAGGTCAATAACAAACTAATGATAATAACAATTGGCTCCAAATAATAACAATTGGTCCCAAA
CTAATTGTATCTTTACAAAGAAGAAATTGAGTGAACAATTCACCCGGTATTTTATTACCTAAACCG

TTTCTCTTGCTGTTATGGTGCGTGAAAGAAGAAATGGGTAAGAAACGGAAATTGACATTTTTGCGT
CAGTGGTGCGTAATGCCCCCATTGTTGGCCAAACACTGATTGATTCGCTGAGGCATCGTGCATACG
CGTCTACCTATGGTAATTTGATGCAGTCTGTCCCATTCTTCCACCAACGCCTGGACAAGTTCATCT
AGCGTGGCTGGTGGCCTTTCACGTTGACGCACACGTCGGCCCAAGATGTCCCAGACATTTTCAATG
GCCAGGGCTCATTGCTGGTCAGGGCATCCTATGGATATTGTGCCGTTGAAGGTGGTTATGTTGTTC
ACATTGAAATTCCAAGTTCTCCTACTCTTTTTAAGAGGAGGTTCACAAAGTACGTTCTTTCATGTT
GGTGAAGAGAATATCAAGGTCTTCTAAGGGATTGTGTCTTATAATATTTGATTTTAAGAAGTTTGA
TATTATCTGCATCCTTCCCAAGAAATTGCAAATGTTCACACACTATTGCGTTTGATAATGTTTTTG
GGGAAATAAACTGTCCAGGACTGCTAAATAGTAATTATTGCTACTTTTAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACTTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTTGACCGCCC
ATGGTGCCAGAGTCGGCCTTCCTTACTGGGACGGGACAACTGCCTTTACAGCTTTGCCAACTTTTG
TCACAGATGAAGAGGACAATCCTTTCCATCAT

INTRON 1G-2/1G-3 (SEQ ID NO:146)

GTGAGTTCACGTAAGCCTACGAGATCAACATTACTCCTTAACAGCCACGGCATCATGTACCGATAT
ATCACAAACAAAAGTATTCAAAGCTTTAAACACGATATGTATGGTTCAAGAATGACATCATTAAAC
AAGGACATGAGTCTGAAATAAACATGACTTGACACCGTTGTGGTCACAGTTTTGTTTCTCATTGGT
GAACCTGTGAAACAACCTTTCAAACCAAAAGATGCCTATTAATATTGTTAATTCCCATGAATTAGG
AGATACACACATTCTACTGTCATTT.....AATAACCGCTTC
CAGCATGAAAACACAATATGATTATCTCAATTCTACCATTACTAATTATAATTTTGACTGGCATTA
TTTGACGACGCGTAAAACATCGCTGCTTTACAGACTGCACTGCGGTAACTGTGACGTTTTCATGAC
GTCACTACATTCTATTCAAAACATTTCCACAGAAGAGCGAGACCACGGCCGTGATGGGTTCTGGGC
AGATGATTACCCAAGTATATATTTATAATAACTTGACTGCTTGCCTGAATAATGTTGACACATGAC
AACGAATTTGTGATAGCGTAAGAAGCGTGAATACTGTGAATAGTGTGAGGGGTGTTTGCTGAGAGT
TAACCACCGTTAATTGCAAAATTCCCGAATACTTGCATTTGCAGTCGAAGAAGAATTGCATTCTTA
CTCCTGTGAATGGACTCATTGTTATTTAGCAGCGGTTATTGAGGTTTTGATCACCTCTAAATAGAC
AATCAGGATGCGGCAAACCGGAAAATTATAGCAGAATCTGTAATTCAAGATGGGCTTGCCTGTGAA
AATATGCTGCGAGTTCAGTAACACTTTTCCCTTTCGATCATGGCCTGTTTTGCTCTGAATCTGGTC
TTTCAGAGGATCCCTGCTTTTTTAAAACTAAAGTCCTCCCAACTCACTTATATTTATGTTTTTTAA
TTATTTATAGTTTTAATATGAACAACAAATCATATTTATTTACACATTATATTTTTCAG

DOMAIN 1G-3 (3rd part of domain g)

GGTCACATAGACTATTTGGGAGTGGATACAACTCGGTCGCCCCGAGACAAGTTGTTCAATGATCCA
GAGCGAGGATCAGAATCGTTCTTCTACAGGCAGGTTCTCTTGGCTTTGGAGCAGACAGAT

Figure 9

Primary structure of the KLH1 protein

DOMAIN B

GLPYWDWTEPMTHIPGLAGNKTYVDSHGASHTNPFHSSVIAFEENAPHTKRQIDQRLFKPATFGHH
TDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYTAFDPLFYFHHSNVDRLWA
VWQALQMRRHKPYRAHCAISLEHMHLKPFAFSSPLNNNEKTHANAMPNKIYDYENVLHYTYEDLTF
GGISLENIEKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKTTDSVQHKAGTFAVLGGSKEMKWGF
DRVFKFDITHVLKDLDLTADGDFEVTVDITEVDGTKLASSLIPHASVIREHARGKLNR

DOMAIN C

VKFDKVPRSRLIRKNVDRLSPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSPEASKKFA
CCVHGMSVFPHWHRLLTVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNP
WFDGHIDTVDKTTTRSVQNKLFEQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGG
AQPYGMASLRYTAFDPLFYLHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANI
NTDHVTKEHSVPFNVFDYKTNFNYEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL
VKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDYEVFDLKPA
SLGKDLFKQPSVIHEPRI

DOMAIN D

GHHEGEVYQAEVTSANRIRKNIENLSLGELESLRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPI
SCCVHGMPTFPHWHRLYVVVVENALLKKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQHHYETN
PFHHGKITHENEITTRDPKDSLFHSDYFYEQVLYALEQDNFCDFEIQLEILHNALHSLLGGKGKYS
MSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNNDEATK
THATPHDGFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRHDRVFAGFLLHNIGTSADGHVFVCL
PTGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQGDFDLEIEITAVNGSHL
DSHVIHSPTILFEAG

DOMAIN E

TDSAHTDDGHTEPVMIRKDITQLDKRQQLSLVKALESMKADHSSDGFQAIASFHALPPLCPSPAAS
KRFACCVHGMATFPQWHRLYTVQFQDSLRKHGAVVGLPYWDWTLPRSELPELLTVSTIHDPETGRD
IPNPFIGSKIEFEGENVHTKRDINRDRLFQGSTKTHHNWFIEQALLALEQTNYCDFEVQFEIMHNG
VHTWVGGKEPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSEANCAVNLMKTPLKP
FSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVRNRQEHSRVFAGFLLEG
FGTSATVDFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTI
TSYDGTVLDSGLIPTPSIIYDPAH

DOMAIN F

HDISSHHLSLNKVRHDLSTLSERDIGSLKYALSSLQADTSADGFAAIASFHGLPAKCNDSHNNEVA
CCIHGMPTFPHWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNIPHLFTDKEYYDVWRNKVMPNP
FARGYVPSHDTYTVRDVQEGLFHLTSTGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGG
PQVYSLSSLHYASYDPIFFIHHSFVDKVWAVWQALQEKRGLPSDRADCAVSLMTQNMRPFHYEINH
NQFTKKHAVPNDVFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSADVQF
QICKTSEDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAV
NGTVLPASILHAPTIIYEPGLG

DOMAIN G

DHHEDHHSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVMADHGPNGYQAIAAFHGNPPMCPMPDGK
NYSCCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDGTTAFTALPTFVTDEEDNPFHHGHID
YLGVDTTRSPRDKLFNDPERGSESFFYRQVLLALEQTD

Figure 10

Genomic sequence of the KLH2 gene

DOMAIN 2B

GGCCTGCCCTACTGGGATTGGACCATGCCAATGAGTCATTTGCCAGAACTGGCTACAAGTGAGACC
TACCTCGATCCAGTTACTGGGGAAACTAAAAACAACCCTTTCCATCACGCCCAAGTGGCGTTTGAA
AATGGTGTAACAAGCAGGAATCCTGATGCCAAACTTTTTATGAAACCAACTTACGGAGACCACACT
TACCTCTTCGACAGCATGATCTACGCATTTGAGCAGGAAGACTTCTGCGACTTTGAAGTCCAATAT
GAGCTCACGCATAATGCAATACATGCATGGGTTGGAGGCAGTGAAAAGTATTCAATGTCTTCTCTT
CACTACACTGCTTTTGATCCTATATTTTACCTCCATCACTCAAATGTTGATCGTCTCTGGGCCATT
TGGCAAGCTCTTCAAATCAGGAGAGGCAAGTCTTACAAGGCCCACTGCGCCTCGTCTCAAGAAAGA
GAACCATTAAAGCCTTTTGCATTCAGTTCCCCACTGAACAACAACGAGAAAACGTACCACAACTCT
GTCCCCACTAACGTTTATGACTATGTGGGAGTTTTGCACTATCGATATGATGACCTTCAGTTTGGC
GGTATGACCATGTCAGAACTTGAGGAATATATTCACAAGCAGACACAACATGATAGAACCTTTGCA
GGATTCTTCCTTTCATATATTGGAACATCAGCAAGCGTAGATATCTTCATCAATCGAGAAGGTCAT
GATAAATACAAAGTGGGAAGTTTTGTAGTACTTGGTGGATCCAAAGAAATGAAATGGGGCTTTGAT
AGAATGTACAAGTATGAGATCACTGAGGCTCTGAAGACGCTGAATGTTGCAGTGGATGATGGGTTC
AGCATTACTGTTGAGATCACCGATGTTGATGGATCTCCCCCATCTGCAGATCTCATTCCACCTCCT
GCTATAATCTTTGACGTGGTCAGAG

INTRON 2B/2C (SEQ ID NO:147)

GTATTTAAAAAAGTAATAAAACCATATTTTCGAATGCGCTTTATGAAATATCGTGTGACTGGTTCT
TTAGTTTACATGGAGTGTAACAACATGCTCCATCAGTTGACATATACTGCTCACACAAAGTAAGGG
ATATTTGATAATGATAACAAATATAATCAAAGCGGTTATACTATCAAGACTTATTCACATAATTAC
AGGTGAAGGGAGGTGTGATCGTGTTCACTGATCAGGTTGAGGCCAGAGAAGTCCCAGTTTGAGTCT
TGCAGAAGATGATGTTTAGGCATGGGGTCGAATCACCAAAATCACATGACTTCAATAACGGGTTGG
ACCACCTCGAGCGACGATGCAAGCAGTAGAGCGTCTACGCATGCTCCTGATAAGGCGACCAATCTG
TTCCTGGGGAATCAGTCGCCACTCCTCTTGTAGTGCCACGCTCATTTCTGCTACGGTCCTGGGTAC
CTGCTATCGGGTCTTGATCCGTATCCCAAGGATGTCCCACACATGTTCAAGGTGAGAGGTCGGGGA
ACATCGCTGGCCACGGTAAGGTCTGAATTTGATGCCGTTGAAAGTGAGCTCTGACAACCTGAGCAT
GGTGAGCTCTGACGTTGTCGTCCTGAAAGATGAATCCAGCTCCATGACAGCGAGCAAAGGGCAGGA
CGTGTTGGTCAATGCAGTTGTCTCTGCAGTACACACCTGTCACTCGCCACTCACAAGCGTGTAGAT
CTGTACGACCAGTCATGGAGATCCCAGCCCACATCATAACGGACCCCTATCCATACCGATCATGAG
CCACCATAGCAGCGTCTTGATGACGTTCTCCCTGTCGCCTCGACATCCTCACACGGCCAAAAGGAA
CGTGGACTCGTCACTGAACATGACATTAGCCAACCTGGCACTTGTCCACCGCTGATGTTGGCGAGA
CCATTCCAGTCGAGCTCTTCGGTGTCTGGCTTTCATCGATAACACGACGTAAGGTCTGCGGGCGTG
CAAGACGGCTCTATGCAGGCGATTTCGGATTGTCTGGGTGCTAACTCTGATCCCAGGTGCCTGCTG
AAGTTGATGCTGGATCTGTGTGGCATTGAGATGGCGATTCCTTAGGACTGTGGAGATGATGAATCG
ATCTTGACTTATGGTGGTGACATTAGGACGTCGGGTTCGTGTCCTATCCTGCACTCTTCCAGTTGT
TCGGTGACGCTCTGGTACCCGGCTGATTACTGACTGAGAATATCCATCTGCCGTGCGACATGAGCC
TGTGTTGGCCCAGCCTGAAGCATTGCAATCGCCAGAGACGCTCTTCAAAAGTCATTCGACGCATGG
TTTTCTGTTCACAAATGACAGCGTAAAACAGTTTTTGGTGCTTTTATGCTTCCCAAGAGCATGAAA
AACACGTTCTATGGGTCGTGCACACCTTACATGACAAGTGTGAAAAGTGACTTGCACCCCCTTGTG
TGTTCGGATGCACACTCTGTTTACGTACTGATGCGATTTGGCGTCTAAACATGTTTTGGCGTCTAA
ACATGTTTTCCTGCATGATTCATATACTATTTTGTCATATTCCTGGCATCAAACCAAACTACAGTG
AAATATATTTCAATATCCCCTACTTTGTGTGAGTAGTATAGATCACTGCAGACAACATATAGACAA
TGCAGTTACACCGTCAACAATCCCAGTCATTAATTATGATGACACTTCCACACATAGTGTCAGTGA
TTGTAATTCAACTGTACACACTTTTCCCGTGAACATTCAGGATCTATATGACTAAATATATAACAT
TAGTATACGTGCAGTTTTGTATCGCTACGACATTGTTGTAACTCTTTGTTTAATCATTTAACAG

DOMAIN 2C

CTGATGCCAAAGACTTTGGCCATAGCAGAAAAATCAGGAAAGCCGTTGATTCTCTGACAGTCGAAG
AACAAACTTCGTTGAGGCGAGCTATGGCAGATCTACAGGACGACAAAACATCAGGGGGTTTCCAGC
AGATTGCAGCATTCCACGGAGAACCAAAATGGTGTCCAAGCCCCGAAGCGGAGAAAAAATTTGCAT
GCTGTGTTCATGGAATGGCTGTTTTCCCTCACTGGCACAGATTGCTGACAGTTCAAGGAGAAAATG
CTCTGAGGAAACATGGATTTACTGGTGGATTGCCCTATTGGGACTGGACTCGGCCAATGAGCGCCC
TTCCACATTTTGTTGCTGATCCTACTTACAATGATTCTGTTTCCAGCCTCGAAGAAGATAACCCAT
GGTATCATGGTCACATAGATTCTGTTGGGCATGATACTACAAGAGCTGTGCGTGATGATCTTTATC
AATCTCCTGGTTTCGGTCACTACACAGATATTGCAAAACAAGTCCTTCTGGCCTTTGAGCAGGACG
ATTTCTGTGATTTTGAGGTACAATTTGAAATTGCCCATAATTTCATACATGCTCTGGTTGGTGGTA
ACGAACCATACAGTATGTCATCTTTGAGGTATACTACATACGATCCAATCTTCTTCTTGCACCGCT
CCAATACAGACCGACTTTGGGCCATTTGGCAAGCTTTGCAAAAATACCGGGGGAAACCATACAACA
CTGCAAACTGTGCCATTGCATCCATGAGAAAACCACTTCAGCCATTTGGTCTTGATAGTGTCATAA
ATCCAGATGACGAAACTCGTGAACATTCGGTTCCTTTCCGAGTCTTCGACTACAAGAACAACTTCG
ACTATGAGTATGAGAGCCTGGCATTTAATGGTCTGTCTATTGCCCAACTGGACCGAGAGTTGCAGA
GAAGAAAGTCACATGACAGAGTCTTTGCAGGATTCCTTCTTCATGAAATTGGAACAGTCTGCACTCG
TGAAATTCTACGTTTGCAAACACAATGTATCTGACTGTGACCATTATGCTGGAGAATTCTACATTT
TGGGAGATGAAGCTGAGATGCCTTGGAGGTATGACCGTGTGTACAAGTACGAGATAACACAGCAGC
TGCACGATTTAGATCTACATGTTGGAGATAATTTCTTCCTTAAATATGAAGCCTTTGATCTGAATG
GCGGAAGTCTTGGTGGAAGTATCTTTTCTCAGCCTTCGGTGATTTTCGAGCCAGCTGCAG

INTRON 2C/2D (SEQ ID NO:148)

GTATGTTTTAAATGTCACTTATCCGTGATCTGTAATGAAGTTAGCAATTCACTTTATCAACTGTTT
GGCTGTACTGTTTCAGTGCGAGTTTTACTTAGGTTGGATTAATTAAAATATTCAAGCTCATAAATG
TTTTGATTCAACTTTTGTTATTTATTTCAAACAG

DOMAIN 2D

GTTCACACCAGGCTGATGAATATCGTGAGGCAGTAACAAGCGCTAGCCACATAAGAAAAAATATCC
GGGACCTCTCAGAGGGAGAAATTGAGAGCATCAGATCTGCTTTCCTCCAAATTCAAAAAGAGGGTA
TATATGAAAACATTGCAAAGTTCCATGGAAAACCAGGACTTTGTGAACATGATGGACATCCTGTTG
CTTGTTGTGTCCATGGCATGCCCACCTTTCCCCACTGGCACAGACTGTACGTTCTTCAGGTGGAGA
ATGCGCTCTTAGAACGAGGGTCTGCAGTTGCTGTTCCTTACTGGGACTGGACCGAGAAAGCTGACT
CTCTGCCATCATTAATCAATGATGCAACTTATTTCAATTCACGATCCCAGACCTTTGATCCTAATC
CTTTCTTCAGGGGACATATTGCCTTCGAGAATGCTGTGACGTCCAGAGATCCTCAGCCAGAACTAT
GGGACAATAAGGACTTCTACGAGAATGTCATGCTGGCTCTTGAGCAAGACAACTTCTGTGACTTTG
AGATTCAGCTTGAGCTGATACACAACGCCCTTCATTCTAGACTTGGAGGAAGGGCTAAATACTCCC
TTTCGTCTCTTGATTATACCGCATTTGATCCTGTATTTTTCCTTCACCATGCAAACGTTGACAGAA
TCTGGGCCATCTGGCAGGACTTGCAGAGATATAGAAAGAAACCATACAATGAGGCTGACTGCGCAG
TCAACGAGATGCGTAAACCTCTTCAACCATTTAATAACCCAGAACTTAACAGTGATTCCATGACGC
TTAAACACAACCTCCCACAAGACAGTTTTGATTATCAAAACCGCTTCAGGTACCAATATGATAACC
TTCAATTTAACCACTTCAGCATACAAAAGCTAGACCAAACTATTCAGGCTAGAAAACAACACGACA
GAGTTTTTGCTGGCTTTATTCTTCACAACATTGGGACATCTGCTGTTGTAGATATTTATATTTGCG
TTGAACAAGGAGGAGAACAAAACTGCAAGACAAAGGCGGGTTCCTTCACGATTCTGGGGGGAGAAA
CAGAAATGCCATTCCACTTTGACCGCTTGTACAAATTTGACATAACGTCTGCTCTGCATAAACTTG
GTGTTCCCTTGGACGGACATGGATTCGACATCAAAGTTGACGTCAGAGCTGTCAATGGATCGCATC
TTGATCAACACATCCTCAACGAACCGAGTCTGCTTTTTGTTCCTGGTGAACGTAAGAATATATATT
ATG

INTRON 2D/2E (SEQ ID NO:149)

GTTATAAAGCAGTATATTCTCTTCAAAAAAGTAGGGGAACTTGGAATTTCAAGGTAAATAACATAA
CTACCTTCAACGGCACAATATCCATATGATGCCCTGGCCAGCAATGAGGCCTGATCTTTTCCCCAT
TAAAAATGTCTGGAACATCTTGGGCAAACGTGTGCGTCAACGTAAAACGCCACCAGTCACGCTAGA
TGAACTTGTCCAGGCGTTGGTGGAAGAATGGGACAGACTGCATCAATTACCATAAGTAGACTCATT
TGCAGCGAATCAGTCAGTGTTTGACCAATAACGGGGGCATTACGCACTACTGACGCAAAACAATGT
CAATTTCCGTTTCTTACCCATTCCTTCTTTCACGGACCATAACAGCAAGAGAAACTGNTTAGGTAA
TGAAATACCGGTGAATTATTGTTAACTGGATTCCTTCTTTGTAAAGATACAATTAGTTTGGGACCA
ATTATTATTATCATTAGTTTGTTATTGACCTTGAAATTCGAAGTTCCTCTACATTTTTTAAGGAGT
TTATTTGATTGACAATGAAATGTAAGAAAAGAGCAAATCGTAAAATACGTTAAAAATTATTCCTTA
AACATCAGTCTCTAACTTCAGTTTAAATTGCCAGTAACACGTGTTATATGATGTTCCGTTTCTCT
TTGTTTTTTAGCATTCAACTTATTTGATATAACGTTTTACTGTTTTAGATTCACATCAAACTGCAG

DOMAIN 2E

ATGGGCTTTCACAACATAATCTTGTGCGAAAAGAAGTAAGCTCTCTTACAACACTGGAGAAACATT
TTTTGAGGAAAGCTCTCAAGAACATGCAAGCAGATGATTCTCCAGACGGATATCAAGCTATTGCTT
CTTTCCACGCTTTGCCTCCTCTTTGTCCAAGTCCATCTGCTGCACATAGACACGCTTGTTGCCTCC
ATGGTATGGCTACCTTCCCTCAGTGGCACAGACTCTACACAGTTCAGTTCGAAGATTCTTTGAAAC
GACATGGTTCTATTGTCGGACTTCCATATTGGGATTGGCTGAAACCGCAGTCTGCACTCCCTGATT
TGGTGACACAGGAGACATACGAGCACCTGTTTTCACACAAAACCTTCCCAAATCCGTTCCTCAAGG
CAAATATAGAATTTGAGGGAGAGGGAGTAACAACAGAGAGGGATGTTGATGCTGAACACCTCTTTG
CAAAAGGAAATCTGGTTTACAACAACTGGTTTTGCAATCAGGCACTATATGCACTAGAACAAGAAA
ATTACTGTGACTTTGAAATACAGTTCGAAATTTTGCATAATGGAATTCATTCATGGGTTGGAGGAT
CAAAGACCCATTCAATAGGTCATCTTCATTACGCATCATACGATCCACTGTTCTATATCCACCATT
CGCAGACAGATCGCATTTGGGCTATCTGGCAAGCTCTCCAGGAGCACAGAGGTCTTTCAGGGAAGG
AAGCACACTGCGCCCTGGAGCAAATGAAAGACCCTCTCAAACCTTTCAGCTTTGGAAGTCCCTATA
ATTTGAACAAACGCACTCAAGAGTTCTCCAAGCCTGAAGACACATTTGATTATCACCGATTCGGGT
ATGAGTATGATTCCCTCGAATTTGTTGGCATGTCTGTTTCAAGTTTACATAACTATATAAAACAAC
AACAGGAAGCTGATAGAGTCTTCGCAGGATTCCTTCTTAAAGGATTTGGACAATCAGCATCCGTAT
CGTTTGATATCTGCAGACCAGACCAGAGTTGCCAAGAAGCTGGATACTTCTCAGTTCTCGGTGGAA
GTTCAGAAATGCCGTGGCAGTTTGACAGGCTTTACAAGTACGACATTACAAAAACGTTGAAAGACA
TGAAACTGCGATACGATGACACATTTACCATCAAGGTTCACATAAAGGATATAGCTGGAGCTGAGT
TGGACAGCGATCTGATTCCAACTCCTTCTGTTCTCCTTGAAGAAGGAAAGC

INTRON 2E/2F (SEQ ID NO:150)

GTATGTATCTCATGTTTCTCAAATAATTTGATTTTCAATGCCCTTACTATAAAGCACAGTTATTGT
TCAGTGCCAGTAACCGTTTATTTACGTAAATGTTACAGGCTATTATAATCAAAAATACATTACCGA
TATTGTTTACCACACAATTATATCATTGTCAAAATCTACCCCCATTACCTGCGTTTTGAATTTGTA
ACCTTCTGACAAAAATGAATTAGCAAGAGCTCTGATGAAGAACATAATGAACAACACCTATCTTTC
TTCTTTCAATGACGGTTTAACAATACAATGCACAATGTAAAAAAATATATATATATATATAATTTT
ATATCTACAGTTAATGCAAATGACTCCACTAATTCAGGGAAACACATTTTCAG

DOMAIN 2F-1 (1st part of domain f)

ATGGGATCAATGTACGTCACGTTGGTCGTAATCGGATTCGTATGGAACTATCTGAACTCACCGAGA
GAGATCTCGCCAGCCTGAAATCTGCAATGAGGTCTCTACAAGCTGACGATGGGGTGAACGGTTATC
AAGCCATTGCATCATTCCACGGTCTCCCGGCTTCTTGTCATGATGATGAGGGACATGAG

INTRON 2F (SEQ ID NO:151)

GTAAAATAAAACGTCCAGTCATCGGAAACCCGCCCAGATATATGGGTTTTTTCTATTTAAACAAA
AAAGCAGAGACAAAAAGATTATTAAAAGTCACATTTAACTTGATATCAGATCAATAGTTTGGCTAG
TTAGTGCTCTATATCCCTCAAATCCTTCGAATCTTTAAGCCTCGTGATATTTTGACAAACAGAGAA
GACTTAGTAGCCCAGACTTTCCCTTATTTTTTCCTGAAAATCTTAATACGGATATTAAATGGATTC
ATTCTGCAACCTACAACCATAGCCCATATGTTATTATTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATTGCCTGTTGTATCCACGGAATGCCAGTATTCCCACACTGGCACAGGCTTTACACCCTGCAAATG
GACATGGCTCTGTTATCTCACGGATCTGCTGTTGCTATTCCATACTGGGACTGGACCAAACCTATC
AGCAAACTGCCTGATCTCTTCACCAGCCCTGAATATTACGATCCTTGGAGGGATGCAGTTGTCAAT
AATCCATTTGCTAAAGGCTACATTAAATCCGAGGACGCTTACACGGTTAGGGATCCTCAGGACATT
TTGTACCACTTGCAGGACGAAACGGGAACATCTGTTTTGTTAGATCAAACTCTTTTAGCCTTAGAG
CAGACAGATTTCTGTGATTTTGAGGTTCAATTTGAGGTCGTCCATAATGCTATTCACTACTTGGTG
GGTGGTCGACAAGTTTATGCTCTTTCTTCTCAACACTATGCTTCATATGACCCAGCCTTCTTTATT
CATCACTCCTTTGTTGACAAAATATGGGCAGTCTGGCAAGCTCTGCAAAAGAAGAGAAAGCGTCCC
TATCATAAAGCGGATTGTGCTCTTAACATGATGACCAAACCAATGCGACCATTTGCACACGATTTC
AATCACAATGGATTCACAAAAATGCACGCAGTCCCCAACACTCTATTTGACTTTCAGGACCTTTTC
TACACGTATGACAACTTAGAAATTGCTGGCATGAATGTTAATCAGTTGGAAGCGGAAATCAACCGG
CGAAAAAGCCAAACAAGAGTCTTTGCCGGGTTCCTTCTACATGGCATTGGAAGATCAGCTGATGTA
CGATTTTGGATTTGCAAGACAGCTGACGACTGCCACGCATCTGGCATGATCTTTATCTTAGGAGGT
TCTAAAGAGATGCACTGGGCCTATGACAGGAACTTTAAATACGACATCACCCAAGCTTTGAAGGCT
CAGTCCATACACCCTGAAGATGTGTTTGACACTGATGCTCCTTTCTTCATTAAAGTGGAGGTCCAT
GGTGTAAACAAGACTGCTCTCCCATCTTCAGCTATCCCAGCACCTACTATAATCTACTCAGCTGGT
GAAG

INTRON 2F-2/2G (SEQ ID NO:152)

GTGAGAGAAACTATAATAGTGTATGTCGGCAAAAAATGTGCTCATATCATGACTCTGTTGGCCGGT
GGTTGCTCTCCTCTCCTCCACCACCACCGGTACCTCCACCTGTCAGGGCATCAATGTACCATG
AAAATGTCTACAATACTAGGCCTCCTGTAGAAGCACGTAAGATTTACATGGCCGGTTTGTAACTAG
TTTAAAGTGCTTCACAGTAACCAAAACCAGTCTCTAAAGATTAATGTCTGTTTAAAATTTAATGCC
ACATTTTCAACTGACATATTCTTGCAATTAAGTACAAATGAAGTAGTATAAATTATCCACAAATAG
CGTGATGCACCACAAATATAAACCGAGTGCTTTTTTGGCATTCCCCACTTGTTCTGGCATGATCAC
ATCATAGATCTCGTTCATGAAGATACTGTTGGATGCTTTTTCCCAATATGCCCCAATCTGTTAAAT
TATTTACACGACCGCAGTGTGTACTTTCATCACTCAGATCTTTACAATGTGTTTGTAACGTTTACA
ATTAGCGTTATGATTGAAATATTACCCCCTGCTACGTTAAATCACATTCACTCACTCATCTGATGT
ACTTTACAGGTCATACCGATGATCACGGCTCAG

DOMAIN 2G-1 (1st part of domain g)

ATCATATTGCTGGCAGTGGAGTCAGGAAAGACGTGACGTCTCTTACCGCATCTGAGATAGAGAACC
TGAGGCATGCTCTGCAAAGCGTGATGGATGATGATGGACCCAATGGATTCCAGGCAATTGCTGCTT
ATCACGGAAGTCCTCCCATGTGTCACATGCCTGATGGTAGAGACGTTGCATGTTGTACTCATG

INTRON 2G-1/2G-2 (SEQ ID NO:153)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAAATGTTTTA
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-2 (2nd part of domain g)

GAATGGCATCTTTCCCTCACTGGCACAGACTGTTTGTGAAACAGATGGAGGATGCACTGGCTGCGC
ATGGAGCTCACATTGGCATACCATACTGGGATTGGACAAGTGCGTTTAGTCATCTGCCTGCCCTAG
TGACTGACCACGAGCACAATCCCTTCCACCAC

INTRON 2G-2/2G-3 (SEQ ID NO:154)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAAATGTTTTA
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-3 (3rd part of domain g)

GGACATATTGCTCATCGGAATGTGGATACATCTCGATCTCCGAGAGACATGCTGTTCAATGACCCC
GAACACGGGTCAGAATCATTCTTCTATAGACAGGTTCTCTTGGCTCTAGAACAGACAGACTTCTGC
CAATTTGAAGTTCAGTTTGAAATAACACACAATGCAATCCACTCTTGGACTGGAGGACATACTCCA
TATGGAATGTCATCACTGGAATATACAGCATATGATCCACTCTTTTATCTCCACCATTCCAACACT
GATCGTATCTGGGCCATCTGGCAGGCACTCCAGAAATACAGAGGTTTTCAATACAACGCAGCTCAT
TGCGATATCCAGGTTCTGAAACAACCTCTTAAACCATTCAGCGAGTCCAGGAATCCAAACCCAGTC
ACCAGAGCCAATTCTAGGGCAGTCGATTCATTTGATTATGAGAGACTCAATTATCAATATGACACA
CTTACCTTCCACGGACATTCTATCTCAGAACTTGATGCCATGCTTCAAGAGAGAAAGAAGGAAGAG
AGAACATTTGCAGCCTTCCTGTTGCACGGATTTGGCGCCAGTGCTGATGTTTCGTTTGATGTCTGC
ACACCTGATGGTCATTGTGCCTTTGCTGGAACCTTCGCGGTACTTGGTGGGGAGCTTGAGATGCCC
TGGTCCTTTGAAAGATTGTTCCGTTACGATATCACAAAGGTTCTCAAGCAGATGAATCTTCACTAT
GATTCTGAGTTCCACTTTGAGTTGAAGATTGTTGGCACAGATGGAACAGAACTGCCATCGGATCGT
ATCAAGAGCCCTACCATTGAACACCATGGAGGAG

INTRON 2G/2H (SEQ ID NO:155)

GTATGTTTTGAGATCCACATAATCTTCTACCCTGTCTCATTTCTAATGCTCTTCAATACACAATTT
ATATAGCCTTTGAGCTTCAGATGTATTACGGACAGGCATTACAGTATACATGTAATATGGTTTTCT
GCTATTTGCAAAAATTGTGTCCTATCTCTGTTCAGATCATCATGGCGGTGACACCTAG

DOMAIN 2H (SEQ ID NO:159)

GTCACGATCACAGTGAACGTCACGATGGATTTTTCAGGAAGGAAGTCGGTTCCCTGTCCCTGGATG
AAGCCAATGACCTTAAAAATGCACTGTACAAGCTGCAGAATGATCAGGGTCCCAATGGATATGAAT
CAATAGCCGGTTACCATGGCTATCCATTCCTCTGCCCTGAACATGGTGAAGACCAGTACGCATGCT
GTGTCCACGGAATGCCTGTATTTCCACATTGGCACAGACTTCATACAATCCAGTTTGAGAGAGCTC
TCAAAGAACATGGTTCTCATTTGGGTCTGCCATACTGGGACTGGAC

Figure 11

Primary structure of the KLH2 protein

DOMAIN B

GLPYWDWTMPMSHLPELATSETYLDPVTGETKNNPFHHAQVAFENGVTSRNPDAKLFMKPTYGDHT
YLFDSMIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSMSSLHYTAFDPIFYLHHSNVDRLWAI
WQALQIRRGKSYKAHCASSQEREPLKPFAFSSPLNNNEKTYHNSVPTNVYDYVGVLHYRYDDLQFG
GMTMSELEEYIHKQTQHDRTFAGFFLSYIGTSASVDIFINREGHDKYKVGSFVVLGGSKEMKWGFD
RMYKYEITEALKTLNVAVDDGFSITVEITDVDGSPPSADLIPPPAIIFDVVR

DOMAIN C

ADAKDFGHSRKIRKAVDSLTVEEQTSLRRAMADLQDDKTSGGFQQIAAFHGEPKWCPSPEAEKKFA
CCVHGMAVFPHWHRLLTVQGENALRKHGFTGGLPYWDWTRPMSALPHFVADPTYNDSVSSLEEDNP
WYHGHIDSVGHDTTRAVRDDLYQSPGFGHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGG
NEPYSMSSLRYTTYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVI
NPDDETREHSVPFRVFDYKNNFDYEYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL
VKFYVCKHNVSDCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKYEAFDLN
GGSLGGSIFSQPSVIFEPAA

DOMAIN D

GSHQADEYREAVTSASHIRKNIRDLSEGEIESIRSAFLQIQKEGIYENIAKFHGKPGLCEHDGHPV
ACCVHGMPTFPHWHRLYVLQVENALLERGSAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPN
PFFRGHIAFENAVTSRDPQPELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYS
LSSLDYTAFDPVFFLHHANVDRIWAIWQDLQRYRKKPYNEADCAVNEMRKPLQPFNNPELNSDSMT
LKHNLPQDSFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFAGFILHNIGTSAVVDIYIC
VEQGGEQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDGHGFDIKVDVRAVNGSH
LDQHILNEPSLLFVPGERKNIYY

DOMAIN E

DGLSQHNLVRKEVSSLTTLEKHFLRKALKNMQADDSPDGYQAIASFHALPPLCPSPSAAHRHACCL
HGMATFPQWHRLYTVQFEDSLKRHGSIVGLPYWDWLKPQSALPDLVTQETYEHLFSHKTFPNPFLK
ANIEFEGEGVTTERDVDAEHLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGG
SKTHSIGHLHYASYDPLFYIHHSQTDRIWAIWQALQEHRGLSGKEAHCALEQMKDPLKPFSFGSPY
NLNKRTQEFSKPEDTFDYHRFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV
SFDICRPDQSCQEAGYFSVLGGSSEMPWQFDRLYKYDITKTLKDMKLRYDDTFTIKVHIKDIAGAE
LDSDLIPTPSVLLEEGK

DOMAIN F

HGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQADDGVNGYQAIASFHGLPASCHDDEGHEIA
CCIHGMPVFPHWHRLYTLQMDMALLSHGSAVAIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNP
FAKGYIKSEDAYTVRDPQDILYHLQDETGTSVLLDQTLLALEQTDFCDFEVQFEVVHNAIHYLVGG
RQVYALSSQHYASYDPAFFIHHSFVDKIWAVWQALQKKRKRPYHKADCALNMMTKPMRPFAHDFNH
NGFTKMHAVPNTLFDFQDLFYTYDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRF
WICKTADDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV
NKTALPSSAIPAPTIIYSAGE

DOMAIN G

DHIAGSGVRKDVTSLTASEIENLRHALQSVMDDDGPNGFQAIAAYHGSPPMCHMPDGRDVACCTHG
MASFPHWHRLFVKQMEDALAAHGAHIGIPYWDWTSAFSHLPALVTDHEHNPPHHGHIAHRNVDTSR
SPRDMLFNDPEHGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYD
PLFYLHHSNTDRIWAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFD

YERLNYQYDTLTFHGHSISELDAMLQERKKEERTFAAFLLHGFGASADVSFDVCTPDGHCAFAGTF
AVLGGELEMPWSFERLFRYDITKVLKQMNLHYDSEFHFELKIVGTDGTELPSDRIKSPTIEHHGG

DOMAIN H (SEQ ID NO:158)

GHDHSERHDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYESIAGYHGYPFLCPEHGEDQYAC
CVHGMPVFPHWHRLHTIQFERALKEHGSHLGLPYWDW